

UNIVERZITA KARLOVA V PRAZE
MATEMATICKO-FYZIKÁLNÍ FAKULTA

Diplomová práce



Ema Bielíková

Intervaly spolehlivosti pro parametr binomického rozdělení

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: prof. RNDr. Jiří Anděl, DrSc.
Studijní program: Matematika
Studijní obor: Pravděpodobnost, matematická
statistika a ekonometrie
Studijní plán: Matematická statistika

Moje poďakovanie na tomto mieste patrí predovšetkým prof. RNDr. Jiřímu Andělovi, DrSc., za pomoc pri písaní diplomovej práce, za cenné rady a pripomienky a najmä za čas, ktorý mi venoval. Ďalej by som sa chcela poďakovať mojej milej rodine za ich trpezlivosť so mnou počas celého štúdia.

Prehlasujem, že som svoju diplomovú prácu napísala samostatne a výhradne s použitím citovaných zdrojov. Súhlasím s požičiavaním práce.

V Prahe, 21. apríla 2006


Ema Bieliková

Název práce: Intervaly spolehlivosti pro parametr binomického rozdělení

Autor: Ema Bieliková

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: prof. RNDr. Jiří Anděl, DrSc.

e-mail vedoucího: Jiri.Andel@mff.cuni.cz

Abstrakt: Konstrukce oboustranných intervalů spolehlivosti pro parametr p binomického rozdělení patří mezi základní problémy řešené v matematické statistice. Často je doporučován interval $\hat{p} \pm \mu_{\alpha/2} n^{-1/2} \sqrt{\hat{p}(1-\hat{p})}$ známy jako standardní Waldův interval. Ukážeme, že pravděpodobnost pokrytí Waldova intervalu je systematicky nižší, a proto uvažujeme nové postupy vedoucí k lepším intervalům spolehlivosti. Pozornost věnujeme zejména čtyřem nejvíce doporučovaným intervalům, a to intervalu založeném na skórovém testu (Wilson(1927)), intervalu navrženém v Agresti a Coull (1998), intervalu založeném na apriorní Jeffreysově hustotě a intervalu konstruovaném pomocí poměru věrohodnosti. Navrhujeme kritérium na posuzování kvality těchto intervalů, které je založené na průměrné pravděpodobnosti pokrytí a na koeficientu střední délky (EV). Edgeworthův rozvoj pravděpodobnosti pokrytí a asymptotický rozvoj střední délky jsou vhodnými nástroji na kompletní studium vlastností těchto intervalů. Výsledky ukazují, že Jeffreysova metoda a metoda založená na věrohodnostním poměru vedou k nejlepším intervalovým odhadům. Pojednání o intervalech spolehlivosti doplňují názorné příklady.
Klíčová slova: Binomické rozdělení, délka intervalu, Edgeworthov rozvoj, koeficient spolehlivosti, intervaly spolehlivosti, pravděpodobnost pokrytí.

Title: Confidence intervals for the parameter of binomial distribution

Author: Ema Bieliková

Department: Department of Probability and Mathematical Statistics

Supervisor: prof. RNDr. Jiří Anděl, DrSc.

Supervisor's e-mail address: Jiri.Andel@mff.cuni.cz

Abstract: In this paper we deal with the classical problem of interval estimation for a binomial proportion. There is a textbook confidence interval for this problem. The interval $\hat{p} \pm \mu_{\alpha/2} n^{-1/2} \sqrt{\hat{p}(1-\hat{p})}$ is known as Wald interval and is currently in near universal acceptance in practice. We first show that the coverage properties of the Wald interval are poor. This leads to consideration of alternative intervals. The four recommended intervals studied in detail are score-test interval (Wilson(1927)), the likelihood-ratio-test interval, a Jeffreys prior Bayesian interval and interval suggested by Agresti and Coull (1998). The numerical tools are proposed here for evaluation of confidence sets, based on average coverage and expected volume. A comparison of these intervals is made through a study of Edgeworth expansion of their coverage probability and a expansion of their expected length. The results suggest that the Jeffreys interval and likelihood ratio interval are the best alternatives to the Wald interval. In this paper we consider the two-side interval estimation problem. The results are complemented by various illustrative examples.

Key words: Binomial distribution, confidence intervals, confidence level, coverage probability, expected volume, Edgeworth expansion.

Obsah

Pár slov úvodom	3
1 Intervaly spoľahlivosti	7
1.1 Formulácia problému	7
1.2 Aplikácia na binomické rozdelenie	10
2 Štandardný interval	13
2.1 Šťastné n , šťastné p	14
2.2 Vlastnosti štandardného intervalu	17
3 Alternatívne intervaly	23
3.1 Wilsonov interval	24
3.1.1 Modifikovaný Wilsonov interval	28
3.2 Agresti-Coullov interval	29
3.3 Jeffreysov interval	31
3.3.1 Modifikovaný Jeffreysov interval	33
3.4 Interval založený na pomere vierohodnosti	33
3.5 Iné alternatívne intervaly	36
3.5.1 Exaktné intervaly spoľahlivosti	36
3.5.2 Mid-P interval	38
3.5.3 Transformácia \hat{p}	40
3.5.4 Modifikácia \hat{p}	41
4 Hodnotenie intervalov spoľahlivosti	42
4.1 Návrh na hodnotenie	43
4.2 Edgeworthov rozvoj	53
4.3 Asymptotická dĺžka intervalov	63
Záverečné poznámky	68

A Dôkazy a odvodenia	71
A.1 Aproximácia kvantilovej funkcia beta rozdelenia	71
A.2 Edgeworthov rozvoj	71
A.3 Asymptotický rozvoj dĺžky	77
Literatúra	82

Pár slov úvodom

V tejto práci sa budeme zaoberať základným, ale veľmi dôležitým problémom v štatistickej praxi. Ide o intervalový odhad pravdepodobnosti p v binomickom rozdelení. Konštruovanie intervalu spoľahlivosti založeného na relatívnej početnosti sa na prvý pohľad zdá byť jednoduchá záležitosť. Mnoho štatistických učebníc budí tento dojem. V súčasnosti je však táto téma veľmi aktuálna a sú jej venované mnohé odborné články.

Poznáme „učebnicový“ interval spoľahlivosti, ktorý má v súčasnosti takmer univerzálnu platnosť. Je známy ako *štandardný interval spoľahlivosti*. Nech náhodná veličina X má binomické rozdelenie $\text{Bi}(n, p)$ s parametrami n a p . Nech n je dané prirodzené číslo a $p \in (0, 1)$ je neznámy parameter. Potom štandardný interval spoľahlivosti je $\hat{p} \pm u_{\alpha/2} n^{-1/2} \sqrt{\hat{p}(1 - \hat{p})}$, kde $\hat{p} = X/n$ je relatívna početnosť a $u_{\alpha/2}$ je $100(1 - \alpha/2)\%$ -ný kvantil normálneho rozdelenia. Uvádza sa tiež ako *Waldov interval*. Interval je názorný, má motivačný charakter. V neposlednom rade je ľahký na počítanie. S výnimkou t-testu, lineárnej regresie a ANOVY, je očividne populárny v každodennej štatistickej praxi. A tak si môžeme myslieť, že daný problém je jednoduchý a má jasné riešenie. V skutočnosti sa v nedávnej minulosti začalo opätovne diskutovať o intervalovom odhade pre parameter binomického rozdelenia. Bolo známe, že pravdepodobnosť pokrytia parametra p pre štandardný interval je nízka pre p blízko 0 alebo 1. Tento fakt viedol k vytvoreniu všeobecného pravidla pre použitie Waldovho intervalu ak $n \cdot \min(p, 1 - p)$ je najmenej 5 (alebo 10). Listovaním v ďalších štatistických textoch môžeme nájsť iné obmedzenia, všetky však poukazujú na slabé vlastnosti štandardného intervalu, len ak p je blízko hraníc a n je malé.

V tejto práci sa budeme zaoberať otázkou spoľahlivosti štandardného intervalu. Napriek rozsiahlej literatúre stále prevláda mylný názor, že problematika štandardného intervalu je vážna, len ak p je blízko 0 alebo 1, alebo ak n je malé. Ukážeme, že v rozpore s všeobecným názorom, problém spoľahlivosti štandardného intervalu pretrváva aj pre celkom veľké npq . Nespoľahlivosť štandardného intervalu prirodzene nastoľuje potrebu hľadania lepších alternatívnych intervalov. Tu „lepšími“ myslíme intervaly, ktorých pravde-

podobnosť pokrytia je bližšie k nominálnej hladine a ktoré sú kratšie. Z praktického hľadiska, ale aj z dôvodu veľkého metodologického významu, potreba vhodného intervalového odhadu je závažná. Hlavným cieľom tejto práce je prezentovanie niekoľkých alternatívnych intervalov a poskytnutie komplexnej štúdie ich vlastností. Pre tento účel budú intervaly podrobené numerickej a teoretickej analýze. Snahou bude v konečnom stanovisku vybrať konkrétnu metódu ako vhodnú alternatívu za štandardný interval pri konštruovaní intervalu spoľahlivosti pre parameter p binomického rozdelenia.

Zvyčajný prístup pri konštruovaní intervalov vychádza z testovania hypotéz. Základné princípy a pojmy konštruovania intervalov spoľahlivosti vo všeobecnom prípade a súvislosti s testovaním hypotéz sú popísané v kapitole 1.

Problematike štandardného intervalu je venovaná kapitola 2. Potom, čo je zavedený pojem *štandardný interval*, prezentujeme niekoľko príkladov, na ktorých môžeme pozorovať slabé vlastnosti intervalu. Vidíme, že štandardný interval je sužovaný osciláciou a negatívnou systematickou odchýlkou v pravdepodobnosti pokrytia. Chyba pochádza z dvoch zdrojov: z diskrétného charakteru a zo šikmosti binomického rozdelenia. V tejto kapitole ide predovšetkým o nastolenie závažnosti a hĺbky problému. Z tohto dôvodu má intuitívny a názorný charakter, pričom teóriu ponechávam na minimálnej úrovni.

Uvedomme si ešte, že chaotické správanie štandardného intervalu sa neodzrkadľuje len v štatistickej teórii, ale má dopad aj pre praxi. Len na ukážku si vezmeme pravdepodobnosť zlyhania v priemyselnej kontrole akosti výrobku. Tu je celkom bežné, že skutočná hodnota p je malá, povedzme p je 0,005. Potom pravdepodobnosť pokrytia parametru p pre 95%-ný interval spoľahlivosti rastie monotónne v n do $n = 592$, kde dosahuje úroveň 0,945, hneď spadne dole na 0,793, keď n je 593. Tieto nízke hodnoty pretrvávajú, až pokiaľ pravdepodobnosť pokrytia vyskočí späť na 0,945 ak $n = 939$. Toto striedanie „šťastných“ a „nešťastných“ úsekov sa deje nepravidelne. A vidíme, že pokrytie parametru p pre štandardný interval môže byť významne menšie i pre celkom veľký výber.

Po prelistovaní 2. kapitoly sa ponúka otázka existencie alternatívnych intervalov, a práve im venujeme ďalšiu pozornosť. V kapitole 3 sú prezentované jednotlivé alternatívne intervaly s ich jasnou motíváciou a pozadím. Za účelom väčšieho dôrazu prezentujeme alternatívne intervaly v dvoch kategóriách: najprv sú prezentované intervaly, ktoré na základe rozsiahlej literatúry a predbežnej analýzy výrazne vynikajú. Tieto intervaly uvádzame ako „odporúčané intervaly“. Podrobnejšie im je venovaná kapitola 4. Osobitne sú prezentované iné alternatívne intervaly, ktoré zmieňujeme práve pre ich motivačný a teoretický charakter, ale ich vlastnosti nie sú tak výrazné. Krátky zoznam odporúčaných intervalov obsahuje interval založený na *Raovom teste*

(score test), interval navrhovaný v Agresti a Coull (1998) [2], interval založený na bayesovskom prístupe a apriórnej *Jeffreysovej hustote* a nakoniec interval vychádzajúci z testu založeného na *vierohodnostnom pomere* (likelihood ratio test). Skórový interval pre binomické rozdelenie po prvýkrát zaviedol Wilson (1927) [25], a tak ho kvôli prehľadnosti označujeme ako *Wilsonov interval*. Agresti a Coull v [2] navrhli interval spoľahlivosti pre špeciálny 95%-ný nominálny koeficient spoľahlivosti. Ďalší z odporúčaných intervalov je tak vhodnou verziou tohoto intervalu pre všeobecný nominálny koeficient spoľahlivosti, a nazýva sa preto *Agresti-Coull*ov interval. Tretí interval, menovite interval odpovedajúci apriórnej *Jeffreysovej hustote*, nazývame *Jeffreysov interval*. Pre štvrtý interval je ponechané terminologické označenie *interval založený na vierohodnostnom pomere*.

V kapitole 4 sú numericky a analyticky porovnávané vlastnosti štandardného intervalu a štyroch odporúčaných intervalov ako sú pravdepodobnosť pokrytia parametra p intervalom a dĺžka intervalu. Pre tento účel definujeme koeficient AC (average coverage), čo je *priemerná pravdepodobnosť pokrytia* a koeficient *strednej dĺžky EV* (expected volume), čo je priemer očakávaných dĺžok intervalov. Priemer berieme cez celý parametrický priestor $(0, 1)$. Uvedené sú podrobné tabuľky AC a EV pre pevné hodnoty parametrov n a $\alpha = 0.05$. Ďalej intervaly študujeme prostredníctvom *Edgeworthovho rozvoja* pravdepodobnosti pokrytia a prostredníctvom asymptotického rozvoja druhého rádu ich strednej dĺžky. Odvodenie Edgeworthovho rozvoja druhého rádu je veľmi pracné a o niečo viac technické, ide však o účinný nástroj na porovnanie a klasifikovanie intervalov s ohľadom k ich pravdepodobnosti pokrytia. Edgeworthov rozvoj vysvetľuje nedostatky štandardného intervalu a umožňuje porovnanie alternatívnych intervalov navzájom. Hlbšie vyšetrenie kompletného rozvoja ukazuje, že *Jeffreysov interval* rovnako ako interval založený na *vierohodnostnom pomere* majú takmer nulové systematické vychýlenie. Taktiež napríklad vidíme, že voľba nominálnej hladiny α môže tak isto ovplyvniť vlastnosti intervalov. V probléme intervalového odhadu pravdepodobnosť pokrytia je len jeden zo spôsobov posudzovania intervalov. Úspornosť (parsimony), prirodzene vyjadrená v zmysle strednej dĺžky intervalu (expected length) je ďalšie podstatné kritérium. Koeficienty u druhého člena asymptotického rozvoja strednej dĺžky intervalu sa pre rôzne intervaly líšia a dávajú tak základ pre ich vzájomné porovnanie. Nakoniec je ešte uvedená integrovaná verzia rozvoja cez celý parametrický priestor $(0, 1)$. Vidíme, že *Agresti-Coull*ov interval je vždy najdlhší, skórový a štandardný interval majú identický rozvoj druhého rádu a *Jeffreysov interval* je vždy najkratší.

Tieto analytické výsledky a dodatočné numerické výpočty ukazujú, že najlepšie zo všetkých uvažovaných alternatív sú *Jeffreysov interval* a interval založený na *vierohodnostnom pomere*, a to v oboch ohľadoch,. Navyše

v pravdepodobnosti pokrytia, aj v strednej dĺžke sú navzájom takmer identické. Skórový interval poskytuje hlavne zlepšenie v pravdepodobnosti pokrytia, ale na úkor jeho strednej dĺžky. Na základe týchto výsledkov môžeme s istotou povedať, že štandardný interval by sme nemali pri konštruovaní intervalu spoľahlivosti v binomickom prípade používať. Môžeme len silno odporučiť používanie jedného alebo viacerých alternatívnych intervalov. Na úrovni základných štatistických kurzov, kde jednoduchosť je tiež jedno významné kritérium, odporúčanie vedie k používaniu Agresti-Coullovho intervalu.

V celej práci uvažujeme konštruovanie jedného obojstranného intervalu spoľahlivosti. Situácia pre jednostranné intervaly sa môže značne líšiť, a to nie len v samotnom prístupe, ale aj v záveroch. Pre jednoduchosť sa obmedzujeme na 95%-ný prípad. Pre ostatné voľby α postupujeme identicky. Pravdepodobnosti, ktoré tu uvádzam, sú výsledkom pravdepodobnostných výpočtov implementovaných v prostredí R. Dodatok (appendix) obsahuje dôkazy a odvodenia.

Kapitola 1

Intervaly spoľahlivosti

V praxi sa stretávame s javmi závisiacimi na hodnote nejakého parametra. Hoci je tento parameter pevný, často je neznámy a musí byť odhadnutý na základe náhodnej vzorky. Avšak odhad parametru je funkcia náhodnej veličiny, tento odhad je sám náhodná veličina a podlieha náhodnej fluktuácii. Konštruovaním intervalov s použitím bodového odhadu parametra a nejakých ďalších predpokladov, nájdeme dve hodnoty, dva koncové body intervalu, ktorý pokrýva skutočnú hodnotu parametru s predom danou spoľahlivosťou, povedzme $1 - \alpha$, kde α leží medzi 0 a 1. Hovoríme o *intervale spoľahlivosti s nominálnym koeficientom spoľahlivosti $1 - \alpha$* . Všeobecne pre celú rodinu diskretných rozdelení platí, že pri intervalovom odhade sa musíme vysporiadať s diskretnou povahou rozdelenia. Preto v diskretnom rozdelení a jednotlivu v binomickom rozdelení obvykle nemôžeme dosiahnuť spoľahlivosť presne rovnú $1 - \alpha$.

1.1 Formulácia problému

Nech náhodná veličina X má rozdelenie s hustotou $f(x; \theta)$ a distribučnou funkciou $F(x; \theta)$, ktorá závisí na jednorozmernom parametre θ z nejakého parametrického priestoru $\Omega \in \mathbb{R}$. Zaoberajme sa konštruovaním *intervalu spoľahlivosti CI* (confidence interval) pre parameter θ . Požadujeme, aby intervalový odhad pokrýval parameter θ s pravdepodobnosťou najmenej $1 - \alpha$ pre všetky hodnoty $\theta \in \Omega$, kde α je predom daná hodnota medzi 0 a 1.

Zvyčajný prístup ku konštruovaniu takýchto *CI* je invertovanie príslušnej rodiny testov: testy *veľkosti* najviac α , každý z nich pre jednu možnú jednoduchú *nulovú hypotézu* $H_0: \theta = \theta_0$ proti obojstrannej *alternatívnej hypotéze* $H_1: \theta \neq \theta_0$. Intervalový odhad dosahuje požadovaný koeficient spoľahlivosti najmenej $1 - \alpha$ práve vtedy, ak každý z odpovedajúcich testov má

dosiahnutú *hladinu testu* najviac α . Pre takýto test a pre každú hodnotu parametru $\theta_0 \in \Omega$ označme $A(\theta_0)$ *obor nezamietnutia* nulovej hypotézy na hladine významnosti α ako doplnok ku kritickému oboru. Ak pozorujeme $X = x$, potom interval spoľahlivosti obsahuje všetky hodnoty parametru θ_0 , pre ktoré nezamietame nulovú hypotézu. Označme

$$CI(x) = \{\theta_0 : x \in A(\theta_0)\}.$$

Dostávame interval spoľahlivosti pre parameter θ

$$CI(x) = [L(x), U(x)],$$

kde $L(x)$, $U(x)$ je ľavý, resp. pravý koncový bod intervalu. Termín *pravdepodobnosť pokrytia* (coverage probability) označuje pravdepodobnosť, že interval spoľahlivosti obsahuje skutočnú hodnotu parametru. Pre pravdepodobnosť pokrytia parametru θ budeme používať značenie

$$C(\theta) = P_\theta(\theta \in CI).$$

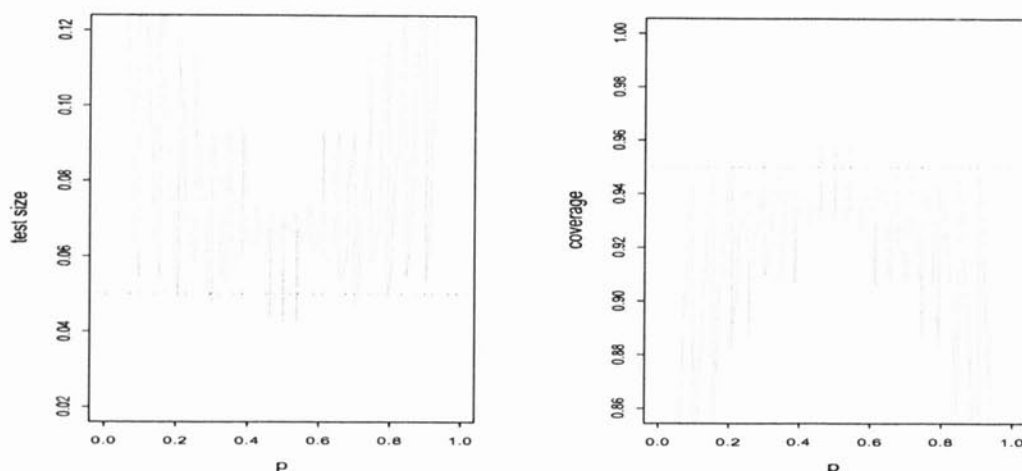
Táto pravdepodobnosť sa mení vzhľadom k hodnote parametru, výraznejšie zmeny sa prejavujú v diskretnom rozdelení. V štatistickej teórii sa definuje *koeficient spoľahlivosti* ako infimum z takýchto pravdepodobností pokrytia cez všetky možné hodnoty parametru $\theta \in \Omega$,

$$C = \inf_{\theta \in \Omega} C(\theta).$$

Hovoríme, že ide o interval spoľahlivosti pre parameter θ s *koeficientom spoľahlivosti* najmenej $1 - \alpha$. Praktici však radšej interpretujú koeficient spoľahlivosti v zmysle „priemerné pokrytie“ ako v zmysle „najhoršie možné pokrytie“.

Práve tak ako inverzia oboru nezamietnutia nulovej hypotézy vedie k intervalom spoľahlivosti, rovnako môžeme prejsť od intervalov spoľahlivosti k testom. V testovaní $H_0: \theta = \theta_0$ proti obojstrannej alternatíve $H_1: \theta \neq \theta_0$, dostaneme test na hladine významnosti α zamietnutím nulovej hypotézy, kedykoľvek hodnota parametru θ_0 nespadá do $100(1 - \alpha)\%$ intervalu spoľahlivosti. Vezmime si nejaký všeobecný test. Test jednoduchej hypotézy proti obojstrannej alternatíve na hladine významnosti α . Za platnosti nulovej hypotézy $\theta = \theta_0$ je dosiahnutá *hladina testu* (*veľkosť testu*), označovaná ako chyba I. druhu, to isté ako hodnota $1 - P_{\theta_0}(\theta_0 \in CI)$, kde druhý člen označuje *pravdepodobnosť pokrytia* intervalu CI parametru θ_0 . Tak ako presnosť intervalového odhadu môžeme vyjadriť prostredníctvom pravdepodobnosti pokrytia pre pevnú hodnotu parametru, presnosť testu udáva *dosiahnutá*

hladina testu. Na obrázku 1.1 vidíme dualitu medzi intervalmi spoľahlivosti a testovaním hypotéz. Ľavý obrázok zobrazuje veľkosť Waldovho asymptotického testu na hladine $\alpha = 0,05$ ako funkciu nulovej hodnoty binomického parametru p_0 . Napravo je pravdepodobnosť pokrytia 95%-ného odpovedajúceho intervalu spoľahlivosti. Preto hovoríme o Waldovom intervale spoľahlivosti. Obrázok 1.1 má len názorný charakter. Bližšie sa o intervaloch spoľahlivosti



Obrázok 1.1: Veľkosť Waldovho testu na hladine $\alpha = 0,05$ a pravdepodobnosť pokrytia odpovedajúceho intervalu v binomickom prípade pre $n = 30$

založených na asymptotických testoch hovorí v kapitole 3. Popri *Waldovom teste* spomenieme *skórový test* (Raov test) a test *pomerom vierohodností*. Už na tomto jednom príklade si môžeme všimnúť, že veľkosť testu ako funkcia parametra θ sa vyznačuje osciláciou, rovnako ako systematickým vychýlením.

U diskretných rozdelení sa stáva že hladina testu nedosahuje predom stanovenú hladinu. Vyplýva to z diskrétnosti povahy rozdelenia. Táto okolnosť vedie k rozšíreniu a doplneniu rodiny testov o *znáhodnené* (alebo *randomizované*) testy. V teórii znáhodnených testov je každý test charakterizovaný funkciou $h(x) : x \rightarrow \langle 0, 1 \rangle$, ktorá sa nazýva *kritická funkcia* testu. Množina všetkých testov sa teda stotožňuje s množinou všetkých funkcií $\{h(x) : 0 \leq h \leq 1\}$. Táto množina je konvexná a kompaktná vzhľadom k slabej konvergencii (dôkaz môžeme nájsť v Lehmann (1986) [19]).

V teórii testovania hypotéz existuje trieda hustôt, pre ktoré sa dajú mnohé úlohy vyriešiť explicitne. Nech náhodná veličina X má hustotu tvaru

$$f(x; \theta) = \exp(\theta.T(x) - A(\theta))u(x),$$

kde $A(\theta)$ je reálna funkcia parametra θ a $T(x)$ a $u(x)$ sú funkcie premennej x . Hovoríme o *jednparametrickej hustote exponenciálneho typu*. Pri riešení optimalizačnej úlohy hľadani najsilnejšieho testu na hladine významnosti α pre $H_0: \theta = \theta_0$ proti $H_1: \theta \neq \theta_0$ sa stretávame s problémom neexistencie rovnomerne najsilnejšieho testu. Preto obmedzíme ponuku testov. Zaoberajme sa testami, ktoré majú hladinu rovnú α a obmedzme sa len na tie, ktoré pre ľubovoľnú alternatívu parametra θ zamietajú H_0 s pravdepodobnosťou najmenej rovnou α . Testy s touto vlastnosťou sa nazývajú *nestranné*. V takejto rodine testov optimálny test skonštruujem na základe nasledovnej vety.

Veta 1.1 *Nech náhodná veličina X má jednparametrickú hustotu exponenciálneho typu. Potom existuje stejnomerne najsilnejší nestranný test hypotézy $H_0: \theta = \theta_0$ proti $H_1: \theta \neq \theta_0$, ktorý má tvar*

$$h(x) = \begin{cases} 1, & \text{ak } T(x) < C_1 \text{ alebo } T(x) > C_2, \\ \gamma_i, & \text{ak } T(x) = C_i, i = 1, 2, \\ 0, & \text{ak } C_1 < T(x) < C_2, \end{cases} \quad (1.1)$$

kde $C_1, C_2, \gamma_1, \gamma_2$ sú volené tak, aby platilo

$$E_{\theta_0} h(x) = \alpha, \quad (1.2)$$

$$E_{\theta_0} [T(x)h(x)] = \alpha E_{\theta_0} [T(x)]. \quad (1.3)$$

Dôkaz: vid' Anděl (2000) [4], strana 174. ■

Rozhodnutie prevedieme nasledovne. Nezamietame hypotézu H_0 vtedy, ak $C_1(\theta_0) < T(X) < C_2(\theta_0)$, ak $X = C_1(\theta_0)$ a $(h(x) < \gamma_1(\theta_0))$, a ak $X = C_2(\theta_0)$ a $(h(x) < \gamma_2)$.

Kompletnosť rodiny testov znamená, že pre každý iný test veľkosti najviac rovné α , existuje taký zástupca z rodiny veľkosti α , ktorý je rovnomerne silnejší proti alternatíve $\theta \neq \theta_0$ — medzi všetkými nestrannými testami (tými, ktoré zamietajú H_0 s väčšou silou pre každú jednu alternatívu). Ak hovoríme o odpovedajúcich intervaloch spoľahlivosti, kompletosť chápeme v zmysle, že pre každý iný *CI* existuje taký interval z uvažovanej rodiny, ktorý má rovnomerne menšiu pravdepodobnosť pokrytia pre každú hodnotu $\theta \neq \theta_0$ — medzi všetkými nestrannými intervalmi spoľahlivosti (tými, pre ktoré pravdepodobnosť pokrytia je vždy najväčšia pre nulovú hypotézu).

1.2 Aplikácia na binomické rozdelenie

Z binomickou premennou sa často stretávame v analýze dat. Nech pozorujeme nejaký jav v postupnosti n nezávislých pokusov. Výskyt tohto javu označu-

jeme ako „úspech“, v opačnom prípade hovoríme o „neúspechu“. Z praktického hľadiska je výhodné použiť numerické premenné: 0 pre neúspech a 1 pre úspech. Takéto náhodné veličiny voláme *Bernoulliho*. Súčet takýchto omnoho jednoduchších náhodných veličín nazývame binomická náhodná veličina, ak sú pozorované javy navzájom nezávislé a pravdepodobnosť úspechu je konštantná.

Nech X je počet úspechov v rade n binomických pokusov s pravdepodobnosťou úspechu $0 < p < 1$, teda X má binomické rozdelenie $X \sim \text{Bi}(n, p)$. V praxi hodnota parametra p je zvyčajne neznáma, a tak ju chceme odhadnúť z náhodného výberu veľkosti n . O parametre p predpokladáme, že nadobúda hodnotu p_0 , a tak testujeme hypotézy $H_0: p = p_0$. Zvyčajne, keď zamietame H_0 , chceme tiež vedieť, či skutočná hodnota parametra je väčšia alebo menšia ako p . Avšak pre prvotné závery je postačujúce rozhodnutie medzi dvoma možnosťami: či data zamietajú alebo nezamietajú nulovú hypotézu. A preto môže byť vhodná formulácia problému ako testu proti obojstrannej alternatíve.

Binomické rozdelenie tvorí jednoparametrický exponenciálny systém. Hustota binomického rozdelenia (vzhľadom k čítacej miere) má tvar

$$f(x; p) = \binom{n}{x} p^x (1-p)^{n-x} = \exp \left\{ x \ln \left(\frac{p}{1-p} \right) + n \ln(1-p) + \ln \binom{n}{x} \right\},$$

čo je jednoparametrická hustota exponenciálneho typu. Pritom $\theta = \ln \frac{p}{1-p}$, $A(\theta) = \ln(1+e^\theta)$ a $T(x) = x$. Rovnomerne najlepší nestranný test je popísaný (1.1), kde $T(x) = x$. Podmienka (1.2) prechádza v tvar

$$\sum_{x=C_1+1}^{C_2-1} \binom{n}{x} p_0^n q_0^{n-x} + \sum_{i=1}^2 (1 - \gamma_i) \binom{n}{C_i} p_0^{C_i} q_0^{n-C_i} = 1 - \alpha.$$

Podmienka (1.3) sa pomocou identity

$$x \binom{n}{x} p_0^n q_0^{n-x} = n p_0 \binom{n-1}{x-1} p_0^{n-1} q_0^{(n-1)-(x-1)}$$

redukuje na

$$\sum_{x=C_1+1}^{C_2-1} \binom{n-1}{x-1} p_0^{x-1} q_0^{(n-1)-(x-1)} + \sum_{i=1}^2 (1 - \gamma_i) \binom{n-1}{C_i-1} p_0^{C_i-1} q_0^{(n-1)-(C_i-1)} = 1 - \alpha.$$

Pri konštruovaní intervalov sa v základe uplatňujú dva prístupy. Prirodzená voľba vedie k intervalom spoľahlivosti s nominálnym koeficientom spoľahlivosti $1 - \alpha$, pre ktoré pravdepodobnosť, že p leží nad a pravdepodobnosť, že p leží pod intervalom spoľahlivosti je približne rovná $\alpha/2$. Koncové body intervalu môžeme explicitne vyjadriť a jednoducho vypočítať s použitím pravdepodobnostných tabuliek alebo vhodného programu. Za druhé, uvažujeme intervaly odpovedajúce rovnomerne najsilnejším testom nulovej hypotézy $p = p_0$ veľkosti najviac α , ktoré majú pravdepodobnosť pokrytia p_0 približne $1 - \alpha$.

V binomickom prípade by bolo výstižnejšie značenie

$$CI = CI(x, n) = [L(x, n), U(x, n)],$$

čím sa zdôrazní, že L a U , ľavý a pravý koniec intervalu, sú výrazy v premennej x , závislé na parametre n . Nech $I(p, x)$ je indikátor rovný 1, ak p patrí do intervalu $CI(n, x)$ a rovný 0 inak. V binomickom rozdelení budeme pre *pravdepodobnosť pokrytia* používať značenie

$$C(p, n) = \sum_{x=0}^n I(p, x) \binom{n}{x} p^x (1-p)^{n-x}.$$

Takýmto indexovaním vyjadrujeme závislosť pravdepodobnosti pokrytia parametra na hodnotách p a n .

Kapitola 2

Štandardný interval

Definujme jeden zo základných problémov riešených v štatistike: konštruovanie intervalov spoľahlivosti CI pre parameter p binomického rozdelenia s koeficientom spoľahlivosti (dosiahnutou pravdepodobnosťou pokrytia parametra p) najmenej $1 - \alpha$, kde α je predom dané. Aj keď pri konštrukcii takýchto intervalov je prirodzené vychádzať priamo z binomického rozdelenia X , za *štandardný interval* považujeme interval založený na normálnej aproximácii. Toto pomenovanie si vyslúžil pre jeho odporúčanie v základných štatistických textoch a pre jeho univerzálne uplatnenie.

Nech je $\phi(x)$ a $\Phi(x)$ hustota, resp. distribučná funkcia štandardného normálneho rozdelenia, tj.

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right), \quad (2.1)$$

$$\Phi(x) = \int_{-\infty}^x \phi(t) dt. \quad (2.2)$$

V súlade s označením $\hat{p} = X/n$, $\hat{q} = 1 - \hat{p}$ a $\kappa = u_{\alpha/2} = \Phi^{-1}(1 - \alpha/2)$ štandardný interval spoľahlivosti CI_S je daný vzorcom

$$CI_S = [\hat{p} - \kappa\sqrt{\hat{p}\hat{q}/n}, \hat{p} + \kappa\sqrt{\hat{p}\hat{q}/n}]. \quad (2.3)$$

Tento interval je inverziou oboru nezamietnutia nulovej hypotézy *Waldovho asymptotického testu*, a preto sa v textoch označuje aj ako *Waldov*. Uvažujem jednorozmerný parametrický priestor a testovanie hypotézy o ľubovoľnom jednorozmernom parametre θ tvaru $H_0: \theta = \theta_0$. Waldov test používa veličinu

$$U_W = \frac{\hat{\theta} - \theta_0}{\widehat{Se}(\hat{\theta})},$$

ktorá má za splnenia určitých predpokladov (viď Anděl (2000) [4], str.149) asymptoticky normálne rozdelenie $N(0, 1)$, kde $\hat{\theta}$ je maximálne virohodný odhad parametra θ a $\widehat{Se}(\hat{\theta})$ je odhad smerodajnej odchýlky $\hat{\theta}$. V binomickom prípade máme $\theta = p$, $\hat{\theta} = X/n$ a $\hat{S}(\hat{\theta}) = \sqrt{\hat{p}\hat{q}/n}$. Obor nezamietnutia nulovej hypotézy je tak daný ako

$$\sqrt{n} \frac{|\hat{p} - p|}{\sqrt{\hat{p}\hat{q}}} \leq u_{\alpha/2}. \quad (2.4)$$

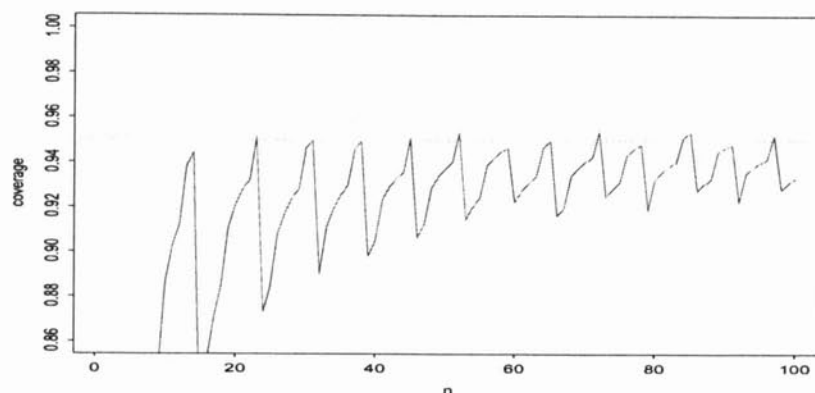
Lahko overíme, že vyriešením nerovnosti (2.4) vzhľadom k p dostaneme (2.3). Je známe, že CI_S garantuje, že $C(n, p) \rightarrow 1 - \alpha$ ako $n \rightarrow \infty$. Preto by sa dalo čakať, že čím je väčšie číslo n , tým je lepšia normálna aproximácia, a tak dosiahnutá pravdepodobnosť pokrytia bude bližšie k nominálnej hladine $1 - \alpha$. Rovnako sú už nejaký čas známe slabé vlastnosti pokrytia tohoto intervalu pre p „blízko“ hraníc. To viedlo k vytvoreniu všeobecne platného obmedzenia pre použitie štandardného intervalu len ak $n \cdot \min(p, 1 - p) \geq 5$. V ďalšom však ukážeme, že tieto domienky sú nesprávne. Poďme sa teda pozrieť, ako CI_S funguje v praxi.

2.1 Šťastné n , šťastné p

Na úvod poznamenajme, že názov podkapitoly je prevzatý z článku Brown, Cai a DasGupta (2001) [7]. Výstižne popisuje obsah tejto časti práce. Neželanou vlastnosťou štandardného intervalu je, že skutočná pravdepodobnosť pokrytia CI_S nezanedbateľne osciluje, a to v obidvoch premenných n i p . Môžeme povedať, že tu existujú nejaké šťastné páry (p, n) , ktorých koeficient spoľahlivosti je blízko k nominálnej hladine. Na druhej strane existujú nejaké „nešťastné“ páry (p, n) , ktorých koeficient spoľahlivosti je značne menší ako $1 - \alpha$. Pritom striedanie týchto úsekov je nepravidelné. Z toho vyplýva, že veľké zmeny v $C(p, n)$ môžeme pozorovať v okolí p pre pevné n a tiež v okolí n pre pevné p . Vieme si predstaviť, že takéto správanie je neželeťné. Uvedieme niekoľko príkladov.

Príklad 2.1

Uvažujme 95%-ný interval spoľahlivosti. Obrázok 2.1 zobrazuje pravdepodobnosť pokrytia ako funkciu rozsahu výberu n pre pevné p , povedzme $p = 0,2$. Naivne sa môžeme domnievať, že pravdepodobnosť pokrytia sa bude s rastúcim n systematicky blížiť k nominálnemu koeficientu. Uvidíme však, že sme ďaleko od pravdy. Napríklad pre máme $C(0,2; 22) = 0,932$, potom $C(0,2; 31) = 0,950$, no pre $C(0,2; 98) = 0,928$, pravdepodobnosť pokrytia je menšia ako pre $n = 22$.



Obrázok 2.1: Štandardný interval; oscilácia v pravdepodobnosti pokrytia pre pevné p

Príklad 2.2

Teraz položíme $p = 0,5$. Pretože $p = 0,5$, očakávame na základe obmedzenia $n \cdot \min(p, 1 - p) \geq 5$ dobré pokrytie už pre n okolo 20. Počítame presnú pravdepodobnosť pokrytia 95%-ného štandardného intervalu pre n od 15 do 50 (čo je malý a stredný výber). V tabuľke 2.1 sú uvedené „šťastné“ a „nešťastné“ n .

Tabuľka 2.1: Šťastné a nešťastné n pre $15 \leq n \leq 65$ a $p = 0.5$

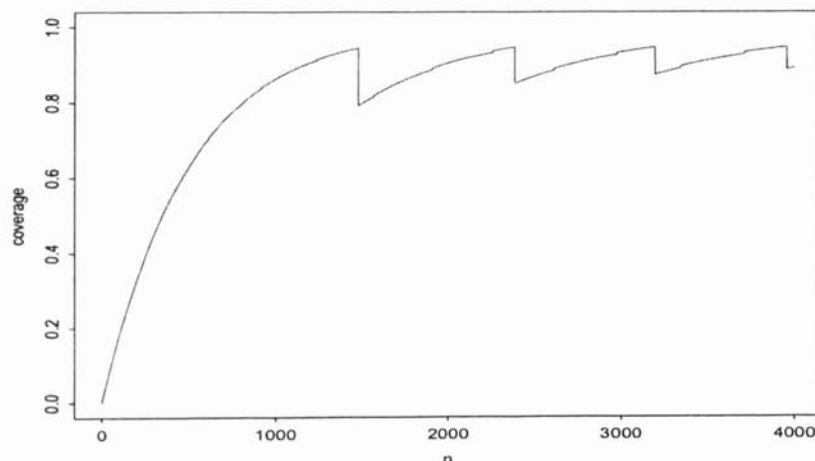
šťastné n	17	20	25	30	35	37	42	44
$C(0,5; n)$	0,951	0,959	0,957	0,57	0,960	0,953	0,956	0,95
šťastné n	49	51	56	58	63	65		
$C(0,5; n)$	0,956	0,951	0,956	0,952	0,957	0,954		
nešťastné n	15	16	18	21	23	26	28	33
$C(0,5; n)$	0,882	0,923	0,904	0,922	0,907	0,925	0,913	0,920
nešťastné n	40	47	54					
$C(0,5; n)$	0,919	0,921	0,924					

Šťastné n definujeme podmienkou $C(p, n) \geq 1 - \alpha$. Takáto silná podmienka pre dosiahnutú pravdepodobnosť pokrytia môže viesť k veľmi nadhodnotenému odhadu. Preto akceptujeme aj určitú odchýlku v pokrytí. Ak predom

dané α chápeme ako mieru presnosti, potom môžeme vychýlenie od nominálnej hladiny vyjadriť cez ľubovoľnú konštantu k predpisom $1 - \alpha - k\alpha$. Tak ako „nešťastné“ označíme tie n , pre ktoré $C(p, n)$ nedosahuje presnosť $k = 1/2$, t.j. hodnotu 0,925. Výsledky odporujú všeobecnej domienke. Hoci p je 0,5, vidíme, že pravdepodobnosť pokrytia je značne malá, dokonca aj pre značne veľké hodnoty n .

Príklad 2.3

Teraz sa presuňme skutočne veľmi blízko hraníc, položíme $p = 0,002$. V tomto prípade neočakávame dobré pokrytie. Z obrázka 2.2 môžeme však pozorovať ďalšiu zvláštnosť. Zo začiatku nepozorujeme osciláciu a pravdepodobnosť pokrytia ma stúpajúcu tendenciu až do $n = 1374$ ($C(p, n) = 0,9507$). Potom však už môžeme pozorovať osciláciu a podobne ako v predchádzajúcich príkladoch, hodnoty n , pre ktoré je pokrytie nedostačujúce sa zjavujú náhodne a nepredvídateľne. V tabuľke 2.2 sú uvedené „nešťastné“ hodnoty n , pri ktorých dochádza k rapidnému a náhlemu poklesu v pravdepodobnosti pokrytia.



Obrázok 2.2: Štandardný interval; Oscilácia v pravdepodobnosti pokrytia pre p blízko nuly

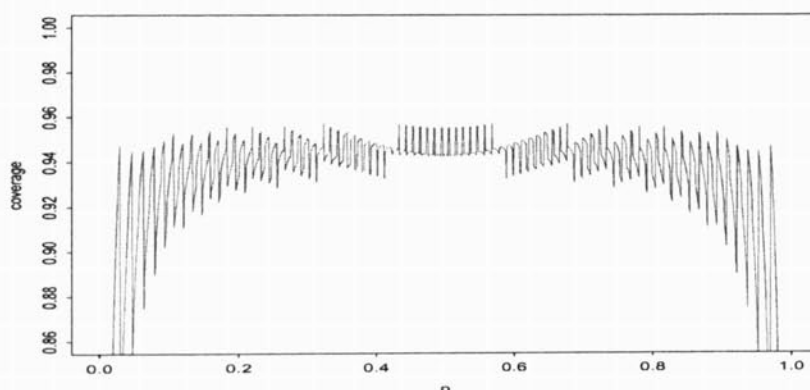
Príklad 2.4

A na záver ešte jeden príklad. Teraz otočme situáciu a uvažujme pravdepodobnosť pokrytia ako funkciu p . Vezmime si nejakú bežnú veľkosť výberu, povedzme $n = 100$. Obrázok 2.3 zobrazuje pravdepodobnosť pokrytia nominálneho 95%-ného štandardného intervalu spoľahlivosti pre $n = 100$ a premennú

Tabuľka 2.2: Náhly pokles v pravdepodobnosti pokrytia pre $p = 0,02$

nešťastné n	1375	2326	3358	3922
$C(0,02; n)$	0,9304	0,9370	0,9387	0,9420

p . Vidíme, že oscilačný efekt skutočne narastá s p blížiacou sa k 0 alebo 1. V týchto oblastiach je značné chaotické správanie štandardného intervalu. Už pre malé zmeny v p , môžeme pozorovať totálne zmeny v pokrytí. S výnimkou hodnôt p v okolí 0,5 je pokrytie z väčšej miery pod nominálnou hladinou. Z uvedených príkladov je evidentné, že skutočná hodnota koeficientu spoľah-

Obrázok 2.3: Štandardný interval; Pravdepodobnosť pokrytia pre pevné $n = 100$ a premennú p

livosti štandardného intervalu sa môže významne líšiť od nominálnej hladiny spoľahlivosti, a to pre stredné, dokonca aj veľké hodnoty n . Rovnaký záver vyplýva aj pre premennú p a nečakané zvraty v pravdepodobnosti pokrytia nastávajú, aj keď p neleží blízko hraníc.

2.2 Vlastnosti štandardného intervalu

Hoci štandardný interval má v súčasnosti univerzálne uplatnenie, ukázali sme si, že jeho vlastnosti sú neakceptovateľné. Špeciálne, slabú pravdepodobnosť pokrytia nepozorujeme iba pre p ležiace blízko 0 alebo 1, ale chaotické správanie v pokrytí pretrváva aj pre veľké, dokonca aj veľmi veľké hodnoty n . Jednoducho môžeme povedať, že štandardný interval spoľahlivosti vykazuje

horšie výsledky, ako by sme si želali. V tejto časti sa snažíme priblížiť zdroj neželateľných javov v pravdepodobnosti pokrytia, a síce *systematického negatívneho vychýlenia* a *oscilácie*. Chyba v pravdepodobnosti pokrytia pochádza z dvoch zdrojov: z *diskrétnosti* a zo *šikmosti* binomického rozdelenia. Ako objasníme v kapitole 4, pre obojstranný interval spoľahlivosti je dominantná chyba, ktorá vyplýva z nespojitosti binomického rozdelenia. A chyba vyplývajúca zo šikmosti je druhotná.

Doteraz sme pozabudli na jednu vec, a síce, že Waldov interval prekračuje rámec parametrického priestoru $(0, 1)$. Ľahkým výpočtom sa dá overiť, že pre prípad $1 - \alpha = 0,95$ a $n = 100$ spodná hranica (2.3) spadá pod nulu pre hodnoty $1 \leq X \leq 3$. To je niečo, čo nemá rozumné zdôvodnenie, a preto interval obmedzíme v 0 alebo v 1. V súlade so zavedením označením, definujeme *Waldov interval spoľahlivosti* ako

$$\left[\frac{X}{n} - \frac{\kappa}{\sqrt{n}} \sqrt{\frac{X}{n} \left(1 - \frac{X}{n}\right)}, \frac{X}{n} + \frac{\kappa}{\sqrt{n}} \sqrt{\frac{X}{n} \left(1 - \frac{X}{n}\right)} \right] \quad (2.5)$$

s výnimkou, že ľavý koniec je 0, ak interval spadá pod 0 a pravý koniec je 1, ak interval prekračuje 1.

Koncové body tohoto intervalu sa dajú veľmi ľahko vypočítať. Všetky výpočty sú prevedené v programe R, Koncové body štandardného intervalu počíta funkcia Wald.

```
Wald <- function(x,n,p,alfa)
{
  emp <- x/n
  emq <- 1-emp
  lower <- max(emp-qnorm(1-alfa/2)*sqrt(emp*emq/n),0)
  upper <- min(emp+qnorm(1-alfa/2)*sqrt(emp*emq/n),1)
  P <- c(lower,upper)
  return(P)
}
```

Štandardný $100(1 - \alpha)\%$ interval spoľahlivosti (2.3) je odvodený z asymptotického *Waldovho testu*. Štandardný interval je množina všetkých takých hodnôt p , ktorých p -hodnota prekračuje α v teste hypotézy $H_0: p = p_0$ proti $H_1: p \neq p_0$ používajúceho testovú štatistiku U_w

$$U_w \equiv \sqrt{n} \frac{X/n - p}{\sqrt{X/n(1 - X/n)}}$$

Štandardný interval je založený na fakte, že *Waldova testová štatistika* U_w má asymptoticky normálne rozdelenie. Avšak dokonca aj pre celkom veľké n skutočné rozdelenie U_w je výrazne nenormálne. Teda už predpoklad, na ktorom je CI_S postavený, je chybný pre mierny, dokonca, aj celkom veľký rozsah výberu n . Pre uistenie sa, uvažujme základné štatistické vlastnosti rozdelenia ako sú *rozptyl, šikmosť a špicatosť*. Pre štandardné normálne rozdelenie nadobúdajú po rade hodnoty 1, 0 a 3. Pre n nie dostatočne veľké odchýlky rozptylu, šikmosti a špicatosti U_w od ich príslušných asymptotických hodnôt sú veľké ako vidíme z tabuľky 2.3. Tabuľka zobrazuje tieto hodnoty pre pevné $p = 0,25$ a n v rozsahu od 40 do 200. Výrazne vysoké hodnoty pozorujeme pre špicatosť.

Tabuľka 2.3: Rozptyl, šikmosť, špicatosť Waldovej náhodnej veličiny

n	40	50	60	70	90	100	150	200
Rozptyl	1,14	1,11	1,08	1,08	1,07	1,07	1,03	1,03
Šikmosť	-0,47	-0,42	-0,35	-0,35	-0,28	-0,26	-0,20	-0,17
Špicatosť	4,00	3,78	3,64	3,57	3,31	3,29	3,16	3,11

Pravdepodobnosť pokrytia parametru p pre štandardný interval spoľahlivosti (2.5) s nominálnym koeficientom spoľahlivosti $1 - \alpha$ je v skutočnosti

$$\begin{aligned}
 & \mathbf{P} \left\{ \left| \frac{X}{n} - p \right| \leq \frac{\kappa}{\sqrt{n}} \sqrt{\left[\frac{X}{n} \left(1 - \frac{X}{n} \right) \right]} \right\} \\
 &= \mathbf{P} \left\{ \left(\frac{X}{n} - p \right)^2 \leq \frac{\kappa^2}{n} \frac{X}{n} \left(1 - \frac{X}{n} \right) \right\} \\
 &= \mathbf{P} \left\{ \left(1 + \frac{\kappa^2}{n} \right) \left(\frac{X}{n} \right)^2 - \left(2p + \frac{\kappa^2}{n} \right) \left(\frac{X}{n} \right) + p^2 \leq 0 \right\}. \quad (2.6)
 \end{aligned}$$

Zavedme označenie $\delta = \sqrt{\kappa^2 + 4np(1-p)}$. Potom riešením kvadratickej nerovnice v X/n môžeme (2.6) prepísať

$$\begin{aligned}
 & \mathbf{P} \left\{ \frac{n(\kappa^2 + 2pn) - \kappa n \delta}{2(\kappa^2 + n)} \leq X \leq \frac{n(\kappa^2 + 2pn) + \kappa n \delta}{2(\kappa^2 + n)} \right\} \\
 &= \mathbf{P} \left\{ \frac{(1-2p)n\kappa^2 - \kappa n \delta}{2(\kappa^2 + n)} \leq X - np \leq \frac{(1-2p)n\kappa^2 + \kappa n \delta}{2(\kappa^2 + n)} \right\}. \quad (2.7)
 \end{aligned}$$

Poznámka 2.5 (Korekcia na spojitosť)

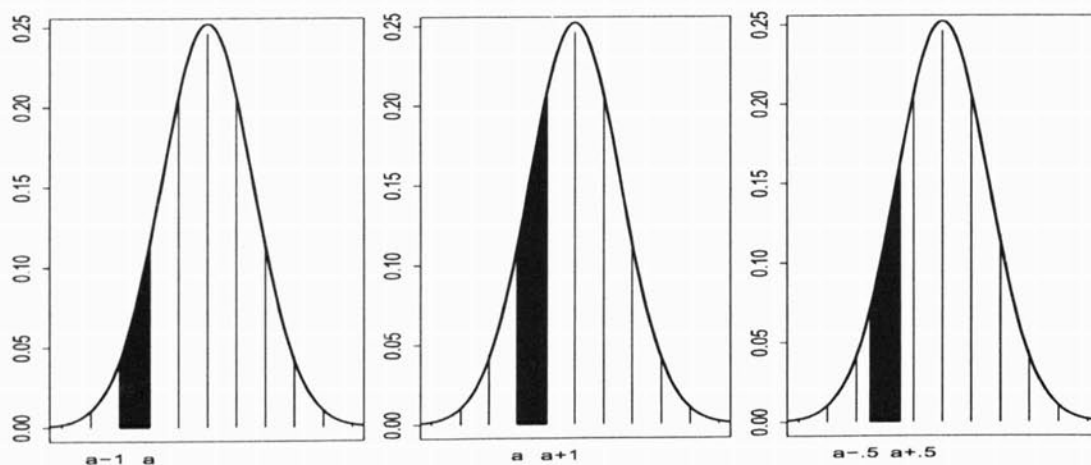
V ďalších úvahách potrebujeme aproximovať rozdelenie binomickej náhodnej veličiny X normálnym rozdelením. Pre tento účel sa obvykle zavádza korekcia na spojitosť. Celý problém spočíva v tom, že normálne rozdelenie môže nadobúdať všetky reálne čísla (je spojité), zatiaľ čo binomické rozdelenie môže nadobúdať iba kladné celé čísla (je diskrétné). Nech X je binomická náhodná veličina s parametrami n a p , potom platí rovnosť

$$P(X \leq a) = P(X < a + 1),$$

pre ľubovoľnú hodnotu $a \in \{0, 1, 2, \dots, n\}$. Túto pravdepodobnosť môžeme aproximovať pravdepodobnosťou

$$P(Y \leq a + 1/2),$$

kde Y je normálne rozdelená náhodná veličina s rovnakou strednou hodnotou a rozptylom ako X , píšeme $E(Y) = np$, $\text{var}(Y) = \sqrt{npq}$. Toto pridanie $1/2$ k x sa nazýva korekcia na spojitosť.



Obrázok 2.4: Korekcia na spojitosť

Obrázok 2.4 nám pomôže lepšie porozumieť, prečo pridávame k binomickej premennej práve $1/2$. V grafoch je vykreslená normálna hustota náhodnej veličiny Y . Zvislé čiary odpovedajú pravdepodobnostiam binomického

rozdelenia v jednotlivých bodoch. Nech X nadobúda hodnotu a . Chceme aproximovať pravdepodobnosť $P(X = a)$ (zvislá čiara v bode a), pravdepodobnosťou, že nastal nejaký jav s normálnym rozdelením (plocha pod krivkou hustoty). Ako vidíme z prvých dvoch obrázkov plocha nad intervalom $(a-1, a)$ je príliš malá a plocha nad $(a, a+1)$ veľká. Lepším priblížením k pravdepodobnosti $P(X = a)$ je obsah plochy pod krivkou hustoty nad intervalom $(a-0.5, a+0.5)$. Nech a a b sú celé čísla z intervalu $\langle 0, n \rangle$ a nech Z je náhodná veličina s rozdelením $N(0, 1)$. Potom korekcia na spojitosť je daná vzorcom

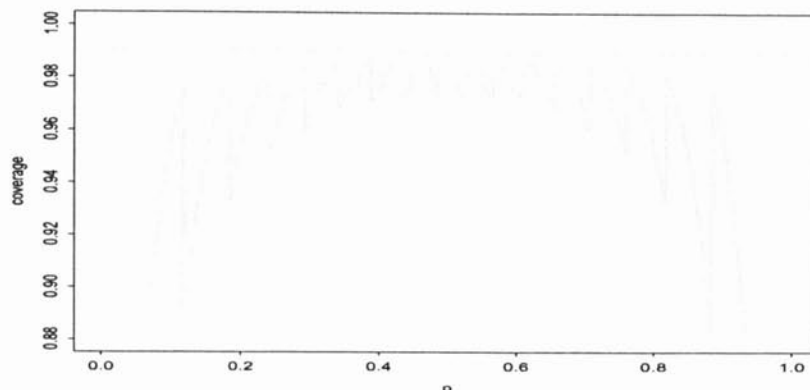
$$P(a \leq X \leq b) \doteq P\left(\frac{a - \frac{1}{2} - np}{\sqrt{np(1-p)}} \leq Z \leq \frac{b + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right).$$

Aplikáciou tohto postupu môžeme pokračovať v odvodzovaní (2.7). Pre zjednodušenie zápisu použijem štandardné označenie $\sigma^2 = np(1-p)$. Máme

$$\begin{aligned} & P\left\{\frac{(1-2p)n\kappa^2 - \kappa n\delta}{2(\kappa^2 + n)} - \frac{1}{2} \leq Y - np \leq \frac{(1-2p)n\kappa^2 + \kappa n\delta}{2(\kappa^2 + n)} + \frac{1}{2}\right\} \\ &= P\left\{\frac{(1-2p)n\kappa^2 - \kappa n\delta}{2(\kappa^2 + n)\sigma} - \frac{1}{2\sigma} \leq Z \leq \frac{(1-2p)n\kappa^2 + \kappa n\delta}{2(\kappa^2 + n)\sigma} + \frac{1}{2\sigma}\right\}. \quad (2.8) \end{aligned}$$

Z tohto prepisu vyplýva, že pre dosiahnutie pravdepodobnosti pokrytia $1-\alpha$, posledný výraz by mal byť tvaru $P\{-\kappa \leq Z \leq \kappa\}$, ale v skutočnosti to tak nie je. Je menší, a to z dvoch dôvodov. Po prvé, násobok kvantilu κ by mal byť 1, ale v skutočnosti je menší ako 1, vždy keď $p(1-p) > 1/(8+4\kappa^2/n)$, čo je pre n veľké splnené vždy, keď $0,15 < p < 0,85$. Po druhé, interval pre Z by mal byť centrován v 0, ale v skutočnosti je stred posunutý k násobku $(1-2p)$. Fakt, že štandardný interval má „zlý“ stred, je príčinou nezanedbateľného systematického záporného vychýlenia pravdepodobnosti pokrytia štandardného intervalu, na ktoré poukazuje predovšetkým nasledujúci príklad. Iba pre tento účel na chvíľu uvažujme 99%-ný nominálny interval spoľahlivosti pre pevné $n = 30$. Na obrázku 2.5 opäť môžeme pozorovať osciláciu, ale čo nás najviac zaráža je fakt, že graf pravdepodobnosti pokrytia nikdy nedosahuje hladinu 0,99. Ďalšie vyhodnocovanie ukazuje, že táto vlastnosť pretrváva až pre n rovné 45. Štandardný interval je centrován v bode $\hat{p} = X/n$. A hoci \hat{p} je maximálne vierohodný a nestranný odhad p , je príčinou *systematického vychýlenia*. Preto, po zanedbaní oscilačného efektu, by sme očakávali, že posunutie stredu intervalu smerom k $1/2$ vedie k zvýšeniu pravdepodobnosti pokrytia. Ako uvidíme neskôr, tento náš dohad je správny.

Oscilácia v pravdepodobnosti pokrytia je spôsobená diskretnou povahou binomického rozdelenia a mnohokrát je demonštrovaná v príkladoch 2.1



Obrázok 2.5: 99% štandardný interval pre $n = 30$ nikdy nedosahuje nominálnu hladinu

až 2.4. S nasledovného výkladu môžeme lepšie porozumieť v intuitívnom smere, prečo pravdepodobnosť pokrytia osciluje. S využitím vzťahu (2.7) vidíme, že malé zmeny v n alebo v p môžu spôsobiť preskočenie koncových bodov intervalu pre X na ďalšiu celú hodnotu. Vezmime si prípad $p = 0,5$ a $\alpha = 0,05$. Keď je $n = 32$ dostávame interval $(11 \leq X \leq 21)$, ale pre $n = 33$ ľavý koniec intervalu preskočí na 12, zatiaľ čo pravý ostáva na 21. A tak množina priaznivých hodnôt pre X stráca bod $X = 11$. Teda $n = 33$ by sme mohli označiť ako tzv. „nešťastné“ n , ak sa vrátíme k spomínanému označeniu.

Kapitola 3

Alternatívne intervaly

Na základe faktov uvedených vyššie je zrejmé, že štandardný interval môžeme považovať za nevhodný. V dôsledku metodologickej potreby takéhoto intervalu stojíme pred závažným problémom hľadania alternatívnych intervalov. Takých, ktoré by lepšie odhadovali parameter p binomického rozdelenia. Avšak neželateľná oscilácia, ktorú sme pozorovali v pravdepodobnosti pokrytia pri štandardnom intervale je spôsobená diskretnou povahou binomického rozdelenia. Oscilácii sa teda nevyhneme ani pri použití alternatívnych procedúr (pokým uvažujeme len nerandomizované procedúry).

V tejto kapitole predstavujeme niekoľko alternatívnych intervalov, každý s ich motiváciou, teoretickým i praktickým významom. Podotýkam, že pri ich výbere sa kvôli širokej ponuke uplatnil aj určitý subjektívny názor. Hlavnou náplňou tejto práce je predstaviť niekoľko takýchto alternatívnych intervalov a na základe analýzy ich vlastností vybrať najvhodnejšiu metódu, ktorú by bolo možné odporučiť ako náhradu za štandardný interval, hoci aj vo vyučovaní procese. Pre tento účel sú v ďalšej kapitole navrhnuté rozhodovacie pravidlá na výber najvhodnejšieho postupu pre konštruovanie intervalov spoľahlivosti. Pri hodnotení alternatívnych intervalov nás zaujímajú dve (resp. tri) hľadiská. Po prvé, pravdepodobnosť pokrytia. Po druhé, úspornosť meraná a chápaná v zmysle dĺžky intervalov. V neposlednom rade sa berie ohľad aj na formálnu jednoduchosť týchto intervalov. Niektorí môžu toto hľadisko dokonca klásť na prvé miesto, pretože intervalový odhad pre parameter p je základný, metodologicky významný problém a intervaly, ktoré majú nepríjemnú formulu nie sú žiadané.

V tomto pojednávaní o alternatívnych intervaloch je uplatnení nasledujúci prístup. Najprv sú prezentované tie intervaly, ktoré na základe čiastočnej analýzy výrazne stoja nad ostatnými kandidátmi a vo svojich vlastnostiach sú vzájomne porovnateľné. Pre tento účel ich označujem spoločným názvom ako „odporúčané intervaly“. Potom osobitne uvádzam niekoľko ďalších in-

tervalov predovšetkým pre ich teoretický a konštruktívny význam. Z nich si špeciálnu pozornosť zaslúži Clopper-Pearsonov interval.

Krátky zoznam „odporúčaných“ intervalov obsahuje skórový interval, interval prezentovaný v Agresti a Coull (1998) [2], interval vychádzajúci z Jeffreysovej apriórnej hustoty pre parameter p a interval konštruovaný na základe testu pomerom vierohodnosti.

3.1 Wilsonov interval

Jednou z alternatív je interval založený rovnako ako štandardná metóda na normálnej aproximácii. Tu však namiesto odhadovanej smerodajnej odchýlky $\sqrt{\hat{p}\hat{q}/n}$ vystupuje nulová smerodajná odchýlka $\sqrt{pq/n}$. Nech $\tilde{X} = X + \kappa^2/2$ a $\tilde{n} = n + \kappa^2$. Klasickým spôsobom označme $\tilde{p} = \tilde{X}/\tilde{n}$ a $\tilde{q} = 1 - \tilde{p}$. Tento interval spoľahlivosti je daný predpisom

$$CI_W = \left[\tilde{p} - \frac{\kappa n^{1/2}}{n + \kappa^2} \sqrt{\hat{p}\hat{q} + \frac{\kappa^2}{4n}}, \tilde{p} + \frac{\kappa n^{1/2}}{n + \kappa^2} \sqrt{\hat{p}\hat{q} + \frac{\kappa^2}{4n}} \right]. \quad (3.1)$$

Interval má nasledovnú teoretickú interpretáciu. Zavedme veličinu

$$U = \frac{X - np}{\sqrt{np(1-p)}},$$

ktorá sa zvykne označovať ako skórová testová štatistika. Veličina U má pre $n \rightarrow \infty$ štandardné normálne rozdelenie $N(0, 1)$. Interval je formovaný na základe testu hypotézy $H_0: p = p_0$, ktorý používa skórovú testovú štatistiku U . Označme $\kappa = u_{\alpha/2}$ ako $100(1 - \alpha/2)\%$ kvantil štandardného normálneho rozdelenia a uvažujeme test veľkosti α . Je daný oborom nezamietnutia nulovej hypotézy $|X - np_0| \leq u_{\alpha/2} \sqrt{np_0(1-p_0)}$. Ide tzv. *skórový test* alebo Raov test. V tejto rodine testov nezamietame nulovú hypotézu vtedy a iba vtedy, ak p_0 leží v intervale CI_W . Skórová testová štatistika, presnejšie jej smerodajná odchýlka vychádza z rozdelenia parametra za platnosti nulovej hypotézy, zatiaľ čo Waldov test vychádza z maximálne vierohodného odhadu parametra. Riešením uvedenej kvadratickej nerovnice v premennej p nájdeme hranice intervalu CI_W

$$\frac{X + \kappa^2/2 \pm \kappa \sqrt{X - X^2/n + \kappa^2/4}}{n + \kappa^2} = \frac{X}{n} + \frac{\kappa^2(1/2 - X/n)}{n + \kappa^2} \pm \frac{\kappa \sqrt{X - X^2/n + \kappa^2/4}}{n + \kappa^2}. \quad (3.2)$$

Je ľahké overiť, že malou úpravou (3.2) dostaneme (3.1). Táto formula nevedie k intuitívnej interpretácii, má však zmysel z praktického hľadiska, a preto uvádzame obidva tvary. V programe R počítame hranice Wilsonovho intervalu funkciou `Wilson`.

```
Wilson <- function(x,n,p,alfa)
{
  emp <- x/n
  emq <- 1-emp
  k <- qnorm(1-alfa/2)
  lower <- (x+(k^2)/2)/(n+k^2)-
    (k*sqrt(n))/(n+k^2)*sqrt(emp*emq+k^2/(4*n))
  upper <- (x+(k^2)/2)/(n+k^2)+
    (k*sqrt(n))/(n+k^2)*sqrt(emp*emq+k^2/(4*n))
  P <- c(lower,upper)
  return(P)
}
```

V literatúre nájdeme interval (3.1) pod názvom *skórový interval spoľahlivosti*. Tento interval po prvýkrát predstavil Wilson (1927) [25], a preto ho budeme označovať *Wilsonov interval*. V [25] môžeme nájsť nasledovnú interpretáciu: nech p je skutočná hodnota parametra a $\sigma = \sqrt{pq/n}$ jeho smerodajná odchýlka. Pozorujme, že hodnota parametra je p_0 . Zavedme pravdepodobnosť P_λ ako pravdepodobnosť, že pozorovaná hodnota parametra p_0 leží mimo intervalu $(p - \lambda\sqrt{pq/n}, p + \lambda\sqrt{pq/n})$. Čebyševova nerovnosť nám dáva hrubý odhad tejto pravdepodobnosti $1/\lambda^2$. Týmto sa dôraz kladie na omylnosť jednotlivých pozorovaní. Toto kritérium môžeme vyjadriť z hľadiska pozorovanej hodnoty p_0 , čo vedie k zostaveniu rovnice $(p_0 - p) = \lambda^2 pq/n$, kde $p = 1 - q$. Rovnica je kvadratická v premennej p a vedie k nájdeniu p . Ak $\lambda^2/n = t$, potom sme našli riešenie

$$p = \frac{p_0 + t/2}{1 + t} \pm \frac{\sqrt{p_0 q_0 t + t^2/4}}{1 + t}.$$

Zmysel je nasledovný: Ak skutočná hodnota p leží mimo tohto rozsahu, pravdepodobnosť, že pozorujeme p_0 je menšia alebo rovná P_λ . (V kontexte celej práce berieme $P_\lambda = \alpha$ a odtiaľ $\lambda = u_{\alpha/2}$.)

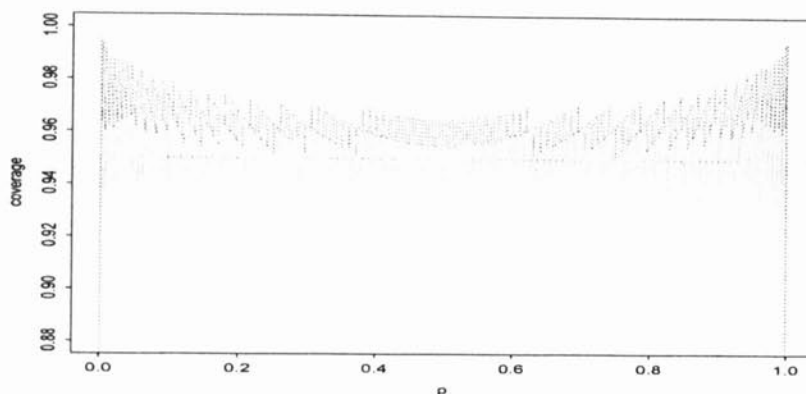
Je informatívne poznamenať, že sa na CI_W môžeme pozeráť ako na určitú modifikáciu štandardného intervalu CI_S . Štandardný interval je centrováný v bode $\hat{p} = X/n$, čo je najlepší nestranný odhad parametra p a používa štandardný odhad $\hat{p}\hat{q}$ rozptylu pq . Wilsonov interval je tak prvý, ktorý posúva stred \hat{p} do

$$\tilde{p} = \frac{X + \frac{1}{2}\kappa^2}{n + \kappa^2} = \frac{n\hat{p} + \frac{1}{2}\kappa^2}{n + \kappa^2} = \hat{p} + \kappa^2 \frac{\frac{1}{2} - \hat{p}}{n + \kappa^2}.$$

Wilsonov interval pridáva úspechom i zlyhaniam po $\frac{1}{2}\kappa^2$ (namiesto zvyčajnej 1). Keďže $|\tilde{p} - \frac{1}{2}| \leq |\hat{p} - \frac{1}{2}|$, táto modifikácia posúva stred intervalu smerom k $\frac{1}{2}$. Ak \hat{p} je blízko 0 alebo 1, mohli by sme si želať iný odhad pre p , a toto je návod. Ďalej máme, že s \tilde{p} považovaným za stred intervalu, Wilsonov interval tiež znižuje rozptyl odhadu $\tilde{p}\tilde{q}$ na

$$\tilde{p}\tilde{q} - \frac{\frac{1}{4}\kappa^2}{n + \kappa^2} = n^2 \frac{\hat{p}\hat{q} + \frac{1}{4n}\kappa^2}{(n + \kappa^2)^2}.$$

Pre intervaly založené na normálnej aproximácii sa zvyčajne navrhuje po-



Obrázok 3.1: Korekcia na spojitosť; pravdepodobnosť pokrytia Wilsonovho intervalu a príslušného korigovaného intervalu (bodkovaná čiara) pre $n = 100$

užitie korekcie na spojitosť (viď poznámku 2.5). Aplikovaním na Wilsonov interval z (3.2) dostávame nasledovné korigované hranice Wilsonovho intervalu

$$\frac{(X \pm \frac{1}{2}) + \kappa^2/2}{n + \kappa^2} \pm \frac{\kappa \sqrt{(X \pm \frac{1}{2}) - (X \pm \frac{1}{2})^2/n + \kappa^2/4}}{n + \kappa^2}, \quad (3.3)$$

s výnimkou, že spodná hranica je 0, ak interval spadá pod 0 a horná hranica je 1, ak interval prekračuje 1.

Tento interval sa zvyčajne používa pre veľké a stredné hodnoty n , rozumieme $n \geq 30$, resp. 100. Zatiaľ, čo odchýlka v pravdepodobnosti pokrytia spôsobená spojitou aproximáciou diskrétného rozdelenia intervalu (3.1) kolíše medzi $\pm P(X = a)$, chyba v (3.3) kolíše medzi 0 a $2P(X = a)$. Teda interval (3.3) garantuje pravdepodobnosť pokrytia najmenej rovnú $1 - \alpha$, zatiaľ čo interval (3.1) pokrýva niektoré hodnoty p s pravdepodobnosťou menšou ako

$1 - \alpha$. Pravdepodobnosť pokrytia u obidvoch intervalov (3.1) a (3.3) môžeme porovnať na obrázku 3.1.

Zaoberajme sa na chvíľu otázkou, aká môže byť hodnota, ktorou pravdepodobnosť pokrytia parametru p pre $p = p_0$ spadne pod požadovanú hodnotu $1 - \alpha$. Nech náhodné veličiny Y a Z sú definované ako v poznámke 2.5. Potom môžeme odchýlku pravdepodobnosti pokrytia od požadovanej hodnoty $1 - \alpha$ vyjadriť ako

$$\begin{aligned} P(X = a) &\doteq P\{|Y - np_0 - \kappa\sqrt{np_0(1-p_0)}| \leq 1/2\} \\ &= P\left\{|Z - \kappa| \leq \frac{1}{2\sqrt{np_0(1-p_0)}}\right\} \\ &\doteq \frac{1}{\sqrt{2\pi}} \exp(-\kappa^2/2) \frac{1}{\sqrt{np_0(1-p_0)}}. \end{aligned}$$

Pre $0,30 \leq p \leq 0,70$ môžeme písať

$$P(X = a) \doteq \frac{2}{\sqrt{2\pi}} \exp(-\kappa^2/2) \frac{1}{\sqrt{n}},$$

Použitím vzťahu uvedeného v Peizer a Pratt (1968) [20], str. 1451 dostaneme

$$\alpha \doteq \frac{2}{\sqrt{2\pi}} \exp(-\kappa^2/2) \cdot \frac{m(\kappa)}{\kappa},$$

kde $m(c)$ je definovaná ako

$$m(c) = 1 - \frac{1}{c^2 + 3 - \frac{1}{0,22(c^2 + 3,2)}}$$

Pomerná časť tejto hodnoty vzhľadom k α môže byť vyjadrená ako

$$\frac{P(X = a)}{\alpha} \doteq \frac{\kappa}{m(\kappa)} \cdot \frac{1}{\sqrt{n}}.$$

Konkrétne pre Wilsonov interval (3.1) a hodnotu $\alpha = 0,05$ máme, že ak za akceptovateľné považujeme prekročenie hodnoty α o najviac 25%, 20%, 10%, 5%, potom interval (3.1) je prijateľný len pre

$$n > 90, 140, 550, 2200 \quad \text{v danom poradí.}$$

3.1.1 Modifikovaný Wilsonov interval

Zaoberajme sa ešte jednou skutočnosťou ohľadom Wilsonovho intervalu. Keďže nejde o tzv. exaktný interval, negarantuje pravdepodobnosť pokrytia najmenej rovnú nominálnemu koeficientu spoľahlivosti. V skutočnosti koeficient spoľahlivosti Wilsonovho intervalu (infimum pravdepodobnosti pokrytia cez všetky hodnoty p) je pod hladinou $1 - \alpha$. Inak povedané, pravdepodobnosť pokrytia Wilsonovho intervalu síce uspokojivo kolíše okolo $1 - \alpha$ pre p nie veľmi blízko hraníc, povedzme $0,30 \leq p \leq 0,70$, ale táto metóda má dve úzke oblasti hodnôt p , jednu blízko 0 a jednu blízko 1, v ktorých skutočná pravdepodobnosť pokrytia spadá výrazne pod nominálnu hladinu spoľahlivosti (a tým celkovo znižuje koeficient spoľahlivosti). Vysvetlenie prečo je tomu tak, spočíva v tom, že existuje oblasť hodnôt parametra p , povedzme $(0, r)$, ktoré spadajú do intervalu (3.1) len pre hodnoty $X = 0$.

Naskytá sa priestor na vylepšenie Wilsonovho intervalu, a to vysporiadaním sa s týmito dvoma úzkymi oblasťami. Uvažujme nasledovne. Pre veľké hodnoty n spodná hranica intervalu spoľahlivosti CI_W pre $X = 1$ je približne $r = (1 + \kappa^2/2 - \kappa\sqrt{1 - \kappa^2/4})/n$. Na druhej strane, je to najväčšia hodnota p , pre ktorú oblasť nazamietnutia hypotézy $p = p_0$ obsahuje práve jeden bod, a to $X = 0$. Pravdepodobnosť pokrytia parametra p takého, že $p < r$ je teda

$$\begin{aligned} P(X = 0) &= [1 - (1 + \kappa^2/2 - \kappa\sqrt{1 - \kappa^2/4})/n]^n \\ &\approx \exp\{-(1 + \kappa^2/2 - \kappa\sqrt{1 - \kappa^2/4})\}. \end{aligned}$$

Pre $n \rightarrow \infty$ a $p_0 = r$ máme $np_0 \rightarrow \lambda = 1 + \kappa^2/2 - \kappa\sqrt{1 - \kappa^2/4}$ a rozdelenie náhodnej veličiny X konverguje k $Po(\lambda)$. Nech T je náhodná veličina s rozdelením $Po(\lambda)$. Pravdepodobnosť pokrytia parametra p pre $p < r$ nahradíme limitnou pravdepodobnosťou $P(T = 0)$. Situácia pre veľké hodnoty p blízko n je analogická.

Táto úvaha vedie ku konštruovaniu intervalu spoľahlivosti, ktorý nazývame *modifikovaný Wilsonov interval* založený na Poissonovej aproximácii a značíme CI_{M-W} . Položme

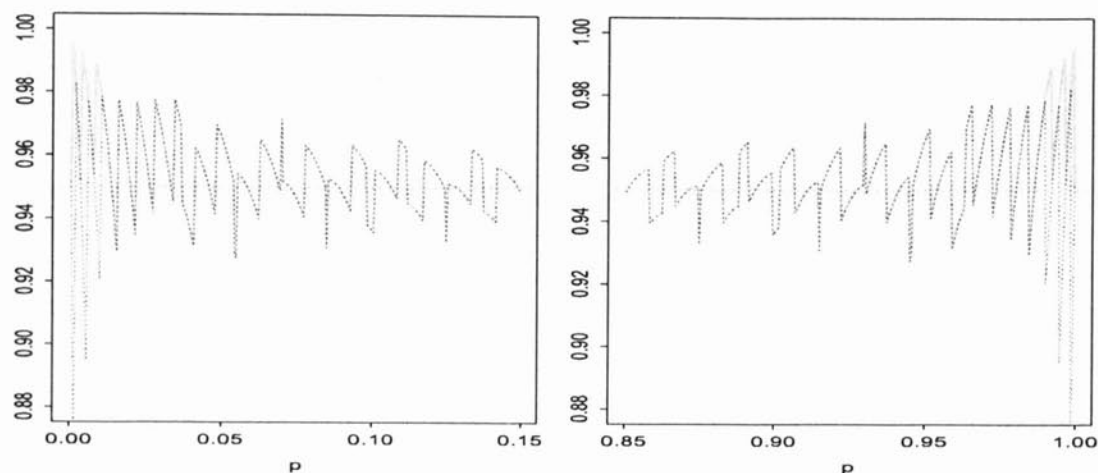
$$x^* = 2 \quad \text{pre } n \leq 50 \quad \text{a} \quad x^* = 3 \quad \text{pre } n > 50.$$

Potom interval CI_{M-W} definujeme predpisom (3.1) s výnimkov modifikovaných dolných koncových bodov intervalu pre $x = 1 \dots x^*$. Pre pevné $1 \leq x \leq x^*$, dolnú hranicu intervalu CI_W nahradíme λ_x/n , kde λ_x je riešením rovnice

$$e^{-\lambda}(\lambda^0/0! + \lambda^1/1! + \dots + \lambda^{x-1}/(x-1)!) = 1 - \alpha. \quad (3.4)$$

Nech $T \sim Po(\lambda)$ a náhodná veličina W_n má χ^2 -rozdelenie s n stupňami voľnosti. Použitím vzťahu medzi Poissonovým rozdelením a χ^2 rozdelením

$$P(T \leq x - 1) = 1 - P(W_{2x} \leq 2\lambda),$$



Obrázok 3.2: Wilsonov interval; pravdepodobnosť pokrytia pre $n = 100$ a $\alpha = 0,05$. Grafy sú symetrické okolo p a pravdepodobnosť pokrytia pre modifikovaný interval (pevná čiara) a pre interval bez modifikácie (bodkovaná) sú inde totožné.

môžeme výrazy λ_x v (3.4) explicitne vyjadriť prostredníctvom χ^2 kvantilov ako

$$\lambda_x = \frac{1}{2} \chi_{2x; \alpha}^2,$$

kde $\chi_{2x; \alpha}^2$ je $100\alpha\%$ -ný kvantil χ^2 -rozdelenia s $2x$ stupňami voľnosti. Symetrický predpis potrebujeme pre modifikovanie hornej hranice pre x veľmi blízko n . Pre pevné $n-x$, kde $1 \leq x \leq x^*$ hornú hranicu Wilsonovho intervalu nahradíme hodnotou $1 - \lambda_x/n$, kde λ_x je riešením rovnice (3.4). Na obrázku 3.2 môžeme porovnať pravdepodobnosť pokrytia obidvoch intervalov.

3.2 Agresti-Coullov interval

Štandardný alebo Waldov interval má veľmi intuitívny tvar jednoduchý na prezentáciu, a to je jeden z dôvodov, prečo je taký rozšírený. Odporúčaný Wilsonov interval má menej názornú formulu (3.1). Preto je potešujúce, že jedna alternatívna metóda má podobný intuitívny predpisom ako CI_S , ale vedie k lepšiemu intervalovému odhadu. Ide o istú zjednodušenú prezentáciu skórovej metódy, a to v nasledujúcom zmysle. Uvažujme konštruovanie 95%-ného intervalu spoľahlivosti. Máme $\alpha = 0,05$ a $u_{\alpha/2}^2 = 1,94^2 \doteq 4$. Potom centrálny

bod Wilsonovho intervalu spoľahlivosti (3.1) je približne

$$\frac{X + u_{\alpha/2}^2/2}{n + u_{\alpha/2}^2} \doteq \frac{X + 2}{n + 4}.$$

Recept na nový interval znie: pridaj dva úspechy a dve zlyhania, a potom použi predpis pre štandardný interval (2.3). Týmto návodom skonštruujeme „vylepšený“ Waldov interval. O tomto intervale pojednáva Agresti a Coull (1998) [2]. Ďaleko sa nelíši od druhého odporúčaného intervalu, ktorý je tu prezentovaný a ktorý z tohto dôvodu označujeme ako *Agresti-Coullov interval* (interval je opäť pomenovaný podľa autorov článku, kde bol popísaný). Nech $\tilde{X} = X + \kappa^2/2$, $\tilde{n} = n + \kappa^2$, $\tilde{p} = \tilde{X}/\tilde{n}$ a $\tilde{q} = 1 - \tilde{p}$. Potom Agresti-Coullov interval spoľahlivosti pre parameter p je definovaný predpisom

$$CI_{AC} = \left[\tilde{p} - \kappa \sqrt{\tilde{p}\tilde{q}/\tilde{n}}, \tilde{p} + \kappa \sqrt{\tilde{p}\tilde{q}/\tilde{n}} \right]. \quad (3.5)$$

Obidva Wilsonov aj Agresti-Coullov interval sú centrované v rovnakej hodnote \tilde{p} . Pritom Agresti-Coullov interval nie je nikdy kratší a vždy obsahuje Wilsonov interval ako vlastný podinterval. Pozastavme sa pri centrálnom bode tohoto intervalu $\tilde{p} = (X + \kappa^2/2)/(n + \kappa^2)$ a považujme ho za bodový odhad parametru p . Rovnako ako pre Wilsonov interval aj tu sa dá ukázať, že tento bodový odhad je silno posunutý smerom k $\frac{1}{2}$ v porovnaní s maximálne vierohodným odhadom $\hat{p} = X/n$. Už Wilson [25] sa zmienil o tomto odhade ako o vhodnej alternatíve k X/n alebo k Laplaceovmu odhadu $(X+1)/(n+2)$. Wilson píše: „Ak rozdelenie šancí na úspech v pozorovanom pokuse nie je symetrické, je možno nespravodlivé zobrať centrálnu hodnotu ako najlepší odhad pravdepodobnosti, ale to je to, čo sa zvyčajne v praxi deje.“ Zdrojový kód tejto metódy v R je nasledovný

```
Agresti <- function(x,n,p,alfa)
{
  k <- qnorm(1-alfa/2)
  emx <- x+(k^2)/2
  emn <- n+k^2
  emp <- emx/emn
  emq <- 1-emp
  lower <- max(emp-k*sqrt(emp*emq/emn),0)
  upper <- min(emp+k*sqrt(emp*emq/emn),1)
  P <- c(lower,upper)
  return(P)
}
```

3.3 Jeffreysov interval

Bayesovské procedúry vychádzajúce z Jeffreysovej apriórnej hustoty majú za sebou široký zoznam dobrých vlastností, a tak je prirodzené uvažovať aj týmto smerom. Podstatný rozdiel medzi týmto a ostatnými prístupmi je, že v baysovskom prístupe považujeme parameter p za náhodnú premennú. Použitím baysovského prístupu vnášame náš predpoklad o skutočnej hodnote parametru p . Beta rozdelenie je štandardne konjugované k binomickému rozdeleniu, a tak je celkom prirodzené použiť k štatistickým záverom o p práve beta hustou.

Nech $X \sim \text{Bi}(n, p)$, $0 < p < 1$. Označme $f(x; p) = \binom{n}{x} p^x (1-p)^{n-x}$ hustotu binomického rozdelenia (vzhľadom k čítacej miere). Fisherovu mieru informácie o parametre p obsiahnutú v X určíme podľa definície ako

$$\begin{aligned} I_n(p) &= \mathbb{E}_p \left[\frac{f'(x; p)}{f(x; p)} \right]^2 = -\mathbb{E}_p \left[\frac{\partial^2 \log f(x; p)}{\partial p^2} \right] \\ &= \mathbb{E}_p \left[\frac{x - np}{p(1-p)} \right]^2 = n \frac{1}{p(1-p)}. \end{aligned}$$

Jeffreysova hustota je až na konštantu rovná $I^{1/2}(p)$. V binomickom rozdelení je Jeffreysova apriórna hustota parametra p $\text{B}(1/2, 1/2)$

$$f(p) = \pi^{-1} p^{-1/2} (1-p)^{-1/2}.$$

Za všeobecného predpokladu, že X má binomické rozdelenie s parametrami n a p a apriórne rozdelenie parametru p je $\text{B}(a, b)$, z Bayesovej vety (viď Anděl (2000) [4]) vyplýva aposteriórne rozdelenie p pre dané X ako $\text{B}(a + X, b + n - X)$.

Tak $100(1 - \alpha)\%$ -ný obojstranný interval vytvorený bayesovským prístupom je daný predpisom

$$[\beta(\alpha/2; X + a, n - X + b), \beta(1 + \alpha/2; X + a, n - X + b)],$$

kde $\beta(\alpha, m_1, m_2)$ označuje α kvantil $\text{B}(m_1, m_2)$ rozdelenia. V bayesovskej analýze má Jeffreysova hustota špeciálne postavenie. Interval pre parameter p binomického rozdelenia založený na Jeffreysovej apriórnej hustote definujeme predpisom

$$CI_J = [L_J(x), U_J(x)],$$

kde $L_J(0) = 0$, $U_J(0) = 1$ a inak

$$\begin{aligned} L_J(x) &= \beta(\alpha/2; X + 1/2, n - X + 1/2), \\ U_J(x) &= \beta(1 - \alpha/2; X + 1/2, n - X + 1/2). \end{aligned} \tag{3.6}$$

Tento interval označujeme ako *Jeffreysov interval*. Je vytvorený zobrazením centrálnych hodnôt p s celkovou pravdepodobnosťou $1 - \alpha$. Takýto interval ponecháva $\alpha/2$ apostriórnej pravdepodobnosti na obidvoch koncoch intervalu. S výnimkou prípadu $x = 0$ (n), kedy je ľavý (pravý) koncový bod intervalu pozmenený. Podobné skrátenie intervalu v hraničných prípadoch uvažujeme pre Waldov interval. Hranice Jeffreysovho intervalu sú $\alpha/2$ a $1 - \alpha$ kvantily $B(x + 1/2, n - x + 1/2)$ rozdelenia a potrebujú byť numericky počítané. Veľmi jednoducho môžeme výsledky získať použitím softvérov ako Minitab, S-Plus, Mathematica a R. V tabuľke uvádzame spodnú a hornú hranicu Jeffreysovho intervalu. Tento interval oproti predchádzajúcim znevýhodňuje

Tabuľka 3.1: Hranice 95% Jeffreysovho intervalu pre $n=30$

x	0	1	2	3	4	5	6	7
spodná	0,00	0,00	0,01	0,03	0,05	0,07	0,09	0,11
horná	0,08	0,14	0,20	0,24	0,29	0,33	0,37	0,40

x	8	9	10	11	12	13	14	15
spodná	0,14	0,16	0,19	0,21	0,24	0,27	0,30	0,33
horná	0,44	0,48	0,51	0,55	0,58	0,61	0,64	0,67

fakt, že bez použitia softvéru nemôžeme získať numerické výsledky priamou cestou. Naskytajú sa dve riešenia. Jedným je použitie tabuliek ako tabuľka 3.1. Druhou možnosťou je počítať približnú hodnotu koncových bodov, a to s použitím normálnych tabuliek. Vhodnú aproximáciu pre tento účel získame z obecnej aproximácie beta kvantilovej funkcie uvedenej v Abramowitz a Stegun (1970) [1], str.945, viď dodatok. Spodná hranica $100(1 - \alpha)\%$ Jeffreysova intervalu je približne

$$\frac{x + 1/2}{n + 1 + (n - x + 1/2)(e^{2\omega} - 1)}, \quad (3.7)$$

kde

$$\omega = \frac{-\kappa\sqrt{4\hat{p}\hat{q}}/n + (\kappa^2 - 3)/(6n^2)}{4\hat{p}\hat{q}} + \frac{(1/2 - \hat{p})(\hat{p}\hat{q}(\kappa^2 + 2) - 1/n)}{6n(\hat{p}\hat{q})^2}.$$

Horná hranica je obdobne s

$$\omega = \frac{\kappa\sqrt{4\hat{p}\hat{q}}/n + \kappa^2 - 3)/(6n^2)}{4\hat{p}\hat{q}} + \frac{(1/2 - \hat{p})(\hat{p}\hat{q}(\kappa^2 + 2) - 1/n)}{6n(\hat{p}\hat{q})^2}.$$

Táto jednoduchá aproximácia je dostatočne presná, avšak pri počítaní tohto intervalu vychádzam z kvantilov beta rozdelenia implementovaných v programe R. Uvediem zdrojový kód metódy Jeffreys.

```
Jeffreys <- function(x,n,p,alfa)
{
  l <- qbeta(alfa/2,x+1/2,n-x+1/2, lower.tail = T)
  u <- qbeta(1-alfa/2,x+1/2,n-x+1/2, lower.tail = T)
  lower <- ifelse(x>0,l,0)
  upper <- ifelse(x<n,u,1)
  P <- c(lower,upper)
  return(P)
}
```

3.3.1 Modifikovaný Jeffreysov interval

Pokrytie parametra p je u Jeffreysovho intervalu pôsobivé a prijateľné, s výnimkou dvoch úzkych oblastí blízko 0 a 1. Na obrázku 4.1 môžeme v týchto oblastiach pozorovať strmé klesajúce piky. Príčina spočíva v tom, že $U_J(0)$ je príliš malé a symetricky $L_J(n)$ je príliš veľké. Nasledovná malá úprava v predpise pre interval sa teda ponúka sama. Položme

$$U_{M-J} = p_1 \quad \text{a} \quad L_{M-J} = 1 - p_1,$$

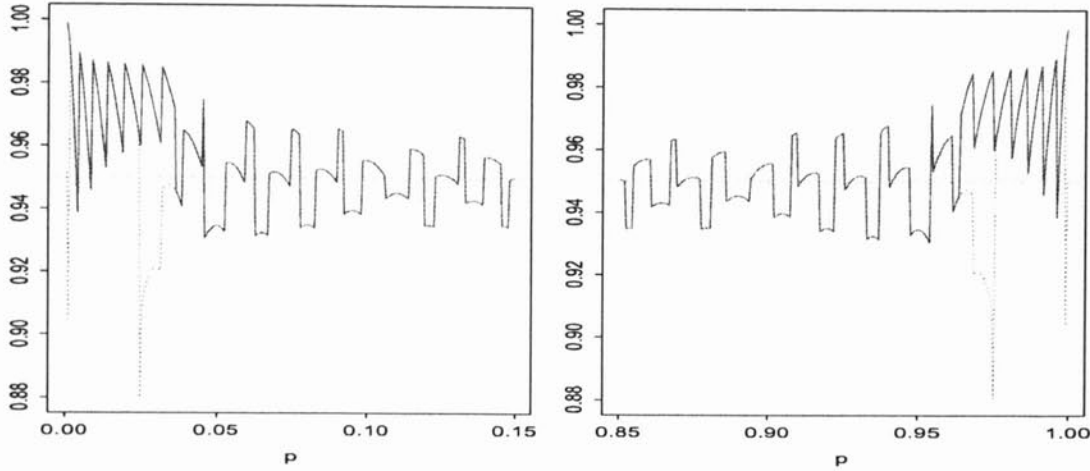
kde p_1 vyhovuje rovnici $(1 - p_1)^n = \alpha/2$ alebo ekvivalentne $p_1 = 1 - (\alpha/2)^{1/n}$. Odtiaľ máme, že $U_J(0)$ resp. $U_J(n)$ sme tak nahradili ich hodnotou vypočítanou z binomickej pravdepodobnosti $P_{p_1}(X = 0) = \alpha/2$. Dodatočne urobíme jednu jednoúčelovú zmenu pre $L_J(1)$ a $U_J(n - 1)$ a položíme

$$L_{M-J}(1) = 0 \quad \text{a} \quad U_{M-J}(n - 1) = 1.$$

V ostatných prípadoch $L_{M-J} = L_J$ a $U_{M-J} = U_J$ zachováваме. Nový interval označujeme CI_{M-J} ako *modifikovaný Jeffreysov interval*. Táto modifikácia odstraňuje dva strmé klesajúce piky v pravdepodobnosti pokrytia Jeffreysovho intervalu a interval tak v tomto smere vylepšuje (viď obrázok 3.3).

3.4 Interval založený na pomere vierohodnosti

Zoznam odporúčaných intervalov uzatvára *interval založený na pomere vierohodnosti*. Metóda maximálnej vierohodnosti je uznávanou metódou v štatistike. Preto konštruovanie takéhoto intervalu má svoje zdôvodnenie. Pretože



Obrázok 3.3: Jeffreysov interval; pravdepodobnosť pokrytia pre $n = 100$ a $\alpha = 0,05$. Grafy sú symetrické okolo p a pravdepodobnosť pokrytia pre modifikovaný interval (pevná čiara) a pre interval bez modifikácie (bodkovaná) sú inde totožné.

princíp maximálnej vierohodnosti nenesie žiadne známe meno, interval sme pomenovali terminologicky a budeme používať označenie CI_{LR} (likelihood ratio).

Tento interval je skonštruovaný na základe testu založenom na vierohodnostnom pomere, ktorý ako testovú štatistiku používa pomer vierohodnosti Λ . Nech $L(\theta)$ označuje vierohodnostnú funkciu. Potom definujeme *pomer vierohodnosti* predpisom

$$\Lambda = \frac{L(\theta_0)}{\sup_{\theta} L(\theta)}, \quad (3.8)$$

kde supremum je brané cez všetky možné hodnoty θ . Tento test vo všeobecnosti nezamieta nulovú hypotézu $H_0: \theta = \theta_0$ ak $-2 \log(\Lambda) \leq \mu_{\alpha/2}^2$. Nech $\hat{\theta}$ je maximálne vierohodný odhadu parametra θ . Zlogaritmovaním (3.8) získame ekvivalentný prepis

$$-2 \ln \Lambda = 2[l(\hat{\theta}) - l(\theta_0)],$$

kde je $l(\theta)$ logaritmická vierohodnostná funkcia. V binomickom prípade dostávame

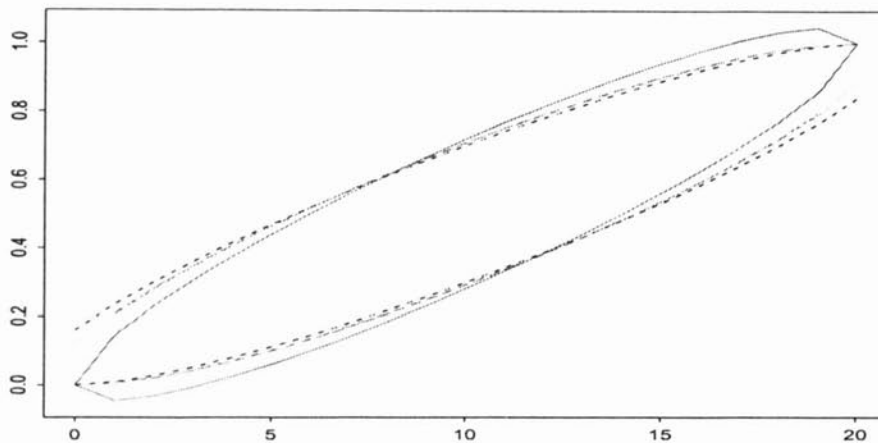
$$\Lambda_n = \frac{p_0^X (1 - p_0)^{n-X}}{(X/n)^X (1 - X/n)^{n-X}}.$$

Určitou nevýhodou tohto intervalu je, že nepoznáme explicitné vyjadrenie jeho koncových bodov v premennej x (a samozrejme n a α) a dokonca pre ne

nemáme ani iný predpis. Avšak sme schopný nájsť ich približné vyjadrenie s presnosťou rádu $n^{-3/2}$. Koncové body intervalu CI_{LR} vyhovujú rovnici $t = p/\hat{p} - 1$, kde t je dané predpisom (viď appendix (A.20) a (A.21)).

$$\begin{aligned}\underline{t} &= -\kappa\sqrt{\hat{q}/\hat{p}}n^{-1/2} + \frac{1}{3}\kappa^2(1-2\hat{p})\frac{1}{\hat{p}}n^{-1} - \frac{1}{36}\kappa^3(1-13\hat{p}\hat{q})\frac{1}{\hat{p}\sqrt{\hat{q}\hat{p}}}n^{-3/2} \\ \bar{t} &= \kappa\sqrt{\hat{q}/\hat{p}}n^{-1/2} + \frac{1}{3}\kappa^2(1-2\hat{p})\frac{1}{\hat{p}}n^{-1} + \frac{1}{36}\kappa^3(1-13\hat{p}\hat{q})\frac{1}{\hat{p}\sqrt{\hat{q}\hat{p}}}n^{-3/2}\end{aligned}$$

a \underline{t} odpovedá spodnému koncovému bodu intervalu a \bar{t} hornému koncovému bodu intervalu.



Obrázok 3.4: Porovnanie hraníc intervalov pre Waldov interval (pevná čiara), Wilsonov (prerušovaná), Agresti-Coull (bodkovaná), Jeffreysov (bodkočiarkovaná) a interval založený na LR (dlhé pomlčky) s $n = 20$

Predstavili sme štyri odlišné intervaly spoľahlivosti ako alternatívy k štandardnému intervalu. Z praktického hľadiska nás zaujíma ako sa líšia jednotlivé intervaly vo svojich koncových bodoch. Ak majú dve odlišné metódy podobné koncové body považujeme ich za totožné. Na obrázku 3.4 sú zobrazené hranice jednotlivých intervalov pre $n = 20$. Vidíme, že Jeffreysov interval a interval založený na vierohodnostnom pomere sú takmer nerozlíšiteľné, Waldov interval je na druhej strane zjavne separovaný. A hranice Wilsonovho intervalu sú jemne nad pre $\hat{p} < 1/2$ a pod opačne. Preto je rozumné očakávať, že obidva, Jeffreysov interval a interval založený na vierohodnostnom pomere, majú porovnateľné vlastnosti. Neskôr v ďalšej časti numerické a teoretické výpočty overia túto domienku.

3.5 Iné alternatívne intervaly

Pozornosť si zaslúžia aj niektoré ďalšie metódy konštruovania intervalov spoľahlivosti. Predovšetkým pre ich teoretický význam. Najmä exaktné intervaly majú v probléme intervalového odhadu pre parameter p binomického rozdelenia osobitný význam. Odvodíme tiež interval získaný transformáciou \hat{p} a spomenieme interval založený na bayesovskom odhade.

3.5.1 Exaktné intervaly spoľahlivosti

V probléme intervalového odhadu parametru p nás trápi slabá pravdepodobnosť pokrytia pre štandardný interval. Na druhej strane, pravdepodobnosť pokrytia takzvaných „exaktných intervalov“ má tendenciu byť o čosi väčšia. A práve im je venovaná táto časť. Bez nich pojednávanie o intervaloch spoľahlivosti, špeciálne o intervaloch spoľahlivosti pre binomické rozdelenie, nemôžeme uzavrieť. Názov *exaktné intervaly* (exact interval) odráža fakt, že sa jedná o intervaly priamo odvodené z binomického rozdelenia. V literatúre sa o nich hovorí ako o „gold standard“, vo význame, že v probléme intervalového odhadu binomického parametra sú akousi zlatou strednou cestou. Pre každú hodnotu parametra p , skutočná pravdepodobnosť pokrytia môže výrazne prekročiť nominálnu hladinu spoľahlivosti, sú zbytočne konzervatívne, čo nie je v štatistickej praxi žetelné, pokiaľ niekto samozrejme striktne nepožaduje pravdepodobnosť pokrytia zdola ohraničenú nominálnym koeficientom spoľahlivosti pre každé p . Tieto intervaly sú v priemere o niečo dlhšie. Niekedy sa však chceme vyhnúť normálnej aproximácii.

Nech $X = x$ a $n \in \mathbb{N}$, potom $100(1 - \alpha)\%$ -exaktný interval spoľahlivosti, označme ho ako $CI_n = [L_n(x), U_n(x)]$, je skonštruovaný na základe binomického testu hypotézy $H_0: p = p_0$ na hladine α . Pre každú jednu hypotézu $p = p_0$ existujú obor nezamietnutia nulovej hypotézy $\{A(p_0) \leq x \leq B(p_0)\}$ na hladine α . Potom pre každé x definujeme

$$CI(x) = \{p_0 : A(p_0) \leq x \leq B(p_0)\}.$$

Za exaktný interval spoľahlivosti považujeme interval splňujúci nasledovné podmienky:

1. *Spojitosť.* Požaduje sa, aby vždy išlo o interval $[L_n(X), U_n(X)]$ daný

$$\{L_n(X) \leq p_0 \leq U_n(x)\} = \{A_n(p_0) \leq x \leq B_n(p_0)\}.$$

2. *Ekvivariancia.* Požadujeme, aby bol problém invariantný voči substitúcii $p \rightarrow 1 - p$ a $x \rightarrow n - x$. Čo znamená, aby táto transformácia

nevedla k žiadným zmenám hodnôt. Píšeme

$$L_n(n-x), U_n(n-x) = 1 - U_n(x), 1 - L_n(x),$$

čo je ekvivalentné k

$$A_n(1-p_0), B_n(1-p_0) = n - B_n(p_0), n - A_n(p_0).$$

3. *Monotónnosť v x*. Pre pevné n je žiadateľné, aby oba koncové intervalové body boli rastúce v x

$$L_n(x+1) > L_n(x) \quad \text{a} \quad U_n(x+1) > U_n(x).$$

4. *Monotónnosť v n*. Pre pevné x je žiadateľné, aby oba koncové intervalové body boli klesajúce v n , čo je

$$L_{n+1}(x) < L_n(x) \quad \text{a} \quad U_{n+1}(x) < U_n(x).$$

Clopper a Pearson(1934) [10] ako prví zverejnili tabuľku binomického intervalu spoľahlivosti v grafickej podobe, a to je tabuľka najčastejšie uvádzaná v textoch, dokonca je priamo implementovaná v softwaroch ako R. *Clopper-Pearsonov interval spoľahlivosti* definujeme predpisom

$$CI_{nCP} = [L_{nCP}(X), U_{nCP}(X)],$$

kde $L_{nCP}(x)$ a $U_{nCP}(x)$ sú riešením rovníc

$$\begin{aligned} P_p(X \geq x) &= \sum_{k=x}^n \binom{n}{k} p^k (1-p)^{n-k} = \alpha/2, \\ P_p(X \leq x) &= \sum_{k=0}^x \binom{n}{k} p^k (1-p)^{n-k} = \alpha/2, \end{aligned}$$

s výnimkou, že ľavý koniec je 0, ak $x = 0$ a pravý koniec je 1, ak $x = n$.

Interval CI_{nCP} teda s výnimkou prípadu $x = 0$ a $x = n$ nemá uzavretú formulu. Označme kde $p_k = P_p(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$, $k = 0 \dots n$. Potom hranica L_{nCP} je minimum z $p \leq x/n$, $p_x + \sum_{k=x+1}^n p_j \geq \alpha/2$ a hranica U_{nCP} je maximum z $p \geq x/n$, $p_x + 1 - \sum_{k=x}^n p_j \geq \alpha/2$. Clopper-Pearsonova interval môžeme vyjadriť predpisom

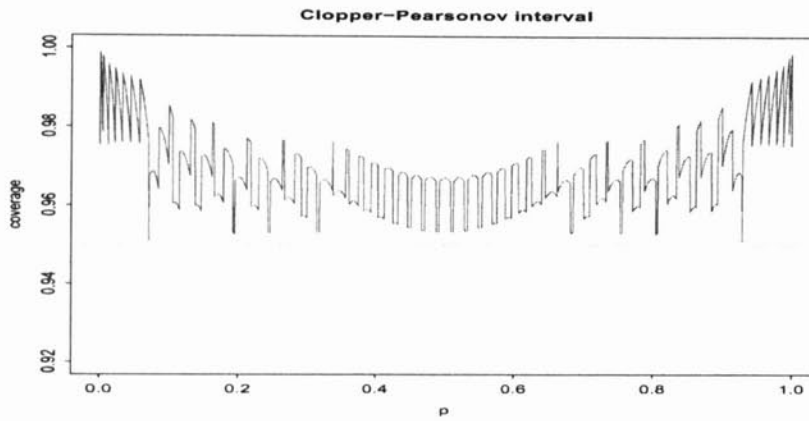
$$CI_{nCP}(X) = [\beta(\alpha/2; x, n-x+1, \beta(1-\alpha/2; x+1, n-x))]. \quad (3.9)$$

Využitím F-rozdelenia môžeme ekvivalentne písať

$$L_{CP}(x) = \left[1 + \frac{n - x + 1}{x F_{2x, 2(n-x+1)}(1 - \alpha/2)} \right]^{-1},$$

$$U_{CP}(x) = \left[1 + \frac{n - x}{(x + 1) F_{2(x+1), 2(n-x)}(\alpha/2)} \right]^{-1},$$

kde $F_{m,n}(c)$ označuje $1 - c$ kvantil F-rozdelenia so stupňami voľnosti m a n .



Obrázok 3.5: Konzervatívnosť; pravdepodobnosť pokrytia pre Clopper-Pearsonov interval

Z vyjadrenie Clopper-Pearsonovho intervalu v (3.9) môžeme vidieť zaujímavú spojitosť medzi Jeffreysovým intervalom a Clopper-Pearsonovým intervalom. Ihneď vidíme, že Jeffreysov interval je vždy obsiahnutý v Clopper-Pearsonovom. V tomto ohľade môžeme CI_J považovať za spojitú verziu ku Clopper-Pearsonovmu. Interval $CI_{n_{CP}}$ používa nerovnosť

$$P\{X \leq A_n(p)\} \leq \alpha/2 \quad \text{a} \quad P\{X \geq B_n(p)\} \leq \alpha/2 \quad (3.10)$$

pre nájdenie dolnej hranice, resp. hornej hranice intervalu. Clopper a Pearson (1934) [10] tak nahradili podmienkou (3.10), klasickú podmienku „koeficientu spoľahlivosti $1 - \alpha$ “, čo je

$$P\{X \leq A_n(p)\} + P\{X \geq B_n(p)\} \leq \alpha.$$

3.5.2 Mid-P interval

Clopper-Pearsonov interval je vo svojich vlastnostiach veľmi konzervatívny ako je zrejmé z obrázka 3.5. Pre elimináciu konzervatívnosti u $CI_{n_{CP}}$ sa

navrhuje namiesto z (3.10) vychádzať z nasledovnej nerovnosti

$$P_p(X \leq A_n(p) - 1) + \frac{1}{2}P_p(X = A_n(p)) = \alpha/2.$$

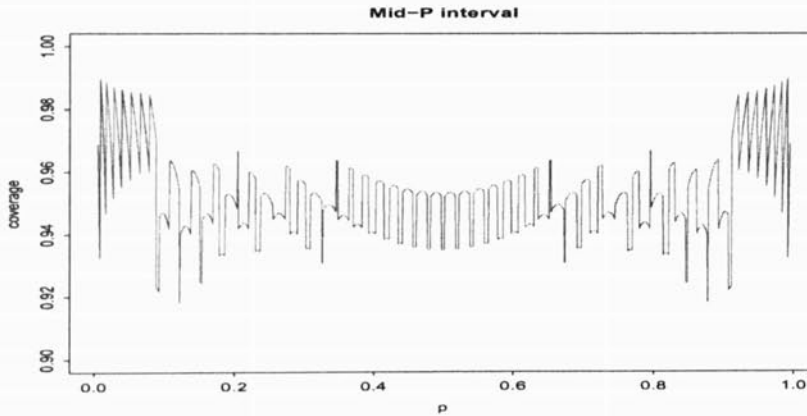
To vedie k riešeniu

$$\begin{aligned} P_{U_{CP}(x)}(X \leq x - 1) + \frac{1}{2}P_{U_{CP}(x)}(X \leq x) = \\ \frac{1}{2}\{P_{U_{CP}(x)}(X \leq x - 1) + P_{U_{CP}(x)}(X \leq x)\} = \alpha/2, \end{aligned}$$

čo môžeme využitím beta rozdelenia vyriešiť pre hornú hranicu

$$U_{mid-P}(x) = \frac{1}{2}\beta(1 - \alpha/2; x, n - x + 1) + \frac{1}{2}\beta(1 - \alpha/2; x + 1, n - x).$$

Symetricky odvodíme spodnú hranicu a dostávame



Obrázok 3.6: Pravdepodobnosť pokrytia pre Mid-P interval

$$L_{mid-P}(x) = (1/2)\beta(\alpha/2; x, n - x + 1) + (1/2)\beta(\alpha/2; x + 1, n - x).$$

Dvoma koncovými bodmi sme definovali nový interval, ktorý sa označuje ako *Mid-P interval*. Platí

$$CI_{mid-P}(X) = [L_{mid-P}(X), U_{mid-P}(X)]. \quad (3.11)$$

Hoci je tento interval značne kratší ako $CI_{n_{CP}}$, stále má uspokojivé pokrytie (viď obrázok 3.6). Ak by sme pokladali x za spojitý argument, neúplná beta funkcia je spojitá a monotónna v x . Potom by sme sa na $U_{mid-P}(x)$

daný v (3.11) mohli pozerat' ako na vážený priemer dvoch neúplných beta funkcií. Pretože priemer dvoch funkcií je približne ich hodnota v strednom bode píšeme

$$U_{mid-P}(x) \approx \beta(1 - \alpha/2; x + 1/2, n - x + 1/2) = CI_J,$$

čo je presne horná hranica Jeffreysovho intervalu.

3.5.3 Transformácia \hat{p}

Vychádzajme zo štandardného intervalu založeného na normálnej aproximácii odhadu $\hat{p} = X/n$. Potom prirodzene uvažujeme o nejakej stabilizujúcej transformácii \hat{p} , označme $g(\cdot)$, ktorá by priniesla lepšiu normálnej aproximáciu. Urobme nasledujúce predpoklady:

- (i) Predpokladajme, že $g(p)$ je spojitá na $[0, 1]$ a má spojitú nenulovú deriváciu v $(0, 1)$. Bez ujmy na všeobecnosti predpokladajme, že $g'(p) > 0$, pre $p \in (0, 1)$.
- (ii) Položme $\sigma^2(p) = (g'(p))^2 pq$ pre $(0 < p < 1)$. Predpokladajme, že $\sigma^2(0) = \lim_{p \rightarrow 0+} \sigma^2(p)$ a $\sigma^2(1) = \lim_{p \rightarrow 1-} \sigma^2(p)$ sú dobre definované a konečné.

Pre $p \in (0, 1)$ máme $\sqrt{n}(g(\hat{p}) - g(p)) = g'(\eta)\sqrt{n}(\hat{p} - p)$, kde $|\eta - p| < |\hat{p} - p|$. Z toho plynie, že $\sqrt{n}(g(\hat{p}) - g(p))/\sigma(\hat{p})$ má v limite rozdelenie $N(0, 1)$. Takto dostávame všeobecný interval spoľahlivosti

$$\left[g^{-1} \left(g(\hat{p}) - \frac{\kappa}{\sqrt{n}} \sigma(\hat{p}) \right), g^{-1} \left(g(\hat{p}) + \frac{\kappa}{\sqrt{n}} \sigma(\hat{p}) \right) \right].$$

Široko rozšírenou stabilizujúcou transformáciou pre binomický prípad je : $g(\hat{p}) = \arcsin(\sqrt{\hat{p}})$. Z vyššie uvedeného dosadením dostávame

$$2\sqrt{n} \left(\arcsin(\sqrt{\hat{p}}) - \arcsin(\sqrt{p}) \right) \rightarrow N(0, 1)$$

Takzvaný *arcsínusový interval* definujeme predpisom pre koncové body intervalu

$$CI_{Arc} = \sin^2(\arcsin(\sqrt{\hat{p}}) \pm \frac{\kappa}{2\sqrt{n}}).$$

Odvodili sme ďalší interval založený na normálnej aproximácii. Prostredníctvom nejakej transformácie môžeme odvodiť aj ďalšie intervaly. Dá sa ale ukázať, viď Chen (1990) [17], že arcsínusový interval je v zmysle dosiahnutého koeficientu spoľahlivosti takmer optimálna transformácia.

3.5.4 Modifikácia \hat{p}

Ďalším zdôvodnením, prečo štandardný interval trpí slabým pokrytím, môže byť fakt, že $\hat{p} = 0$ alebo $\hat{p} = 1$ kedykoľvek $X = 0$ alebo $X = n$. Tieto odhady sú extrémne a prekračujú parametrický priestor. A tak je prirodzené pozeráť sa po inom odhade pre p , ktorý by substituoval \hat{p} v predpise (2.4) pre štandardný interval. Alternatívne riešenie poskytuje bayesovský odhad. Nech p má apriórne $B(a, b)$ rozdelenie, bayesovský odhad pre parameter p je $\tilde{p} = (X + a)/(n + a + b)$. Z hľadiska invariance je výhodnejšie zobrať $a = b$, tak dostávame bayesovský odhad $(X + a)/(n + 2a)$, ktorý je pre extrémne hodnoty X rozumnejším odhadom ako \hat{p} . Na základe tejto úvahy sme dospeli k intervalu spoľahlivosti pre parameter, ktorý označíme CI_a a je určený predpisom

$$CI_a = \frac{X + a}{n + 2a} \pm \frac{\kappa}{\sqrt{n}} \left(\frac{X + a}{n + 2a} \left(1 - \frac{X + a}{n + 2a} \right) \right).$$

Vlastnosti tohoto intervalu závisia na voľbe a . Pre voľbu $a = \kappa^2/2$ hovoríme o Agresti-Coullom intervale. Jeho vlastnosti bližšie rozoberáme v nasledujúcej kapitole.

Kapitola 4

Hodnotenie intervalov spoľahlivosti

Konštruovanie intervalov spoľahlivosti pre parameter p binomického rozdelenia sa v tejto dobe radí medzi aktuálne témy. Študovaním v súčasnosti už rozšírených textov na túto tému a následnou vlastnou predbežnou analýzou alternatívnych intervalov, prevedenou v programe R, sme intervaly spomínané v predchádzajúcej kapitole označili ako „odporúčané“. V snahe v probléme intervalového odhadu parametra p binomického rozdelenie dospieť k nejakému záveru a nájsť jednu alebo dve alternatívne metódy konštruovania intervalov spoľahlivosti vhodné pre praktické použitie, v tejto kapitole porovnávame štyri odporúčané intervaly a navrhujeme vhodné rozhodovacie nástroje.

Uvažujme všeobecný interval spoľahlivosti $CI = [L(X), U(X)]$, určený dvoma koncovými bodmi intervalu $L(X)$, resp. $U(X)$. Základné hodnotenie intervalu spoľahlivosti vychádza z definície pravdepodobnosti pokrytia

$$P_p(p \in CI) = P_p\{L(X) \leq p \leq U(X)\} = 1 - \alpha,$$

čím požadujeme, aby interval spoľahlivosti CI pokrýval skutočnú hodnotu p s pravdepodobnosťou $1 - \alpha$. Osobitne v binomickom prípade nie je obvykle možné nájsť také koncové body intervalu $U(X)$ a $L(X)$, ktorého spoľahlivosť by sa presne rovnala nominálnemu koeficientu spoľahlivosti. Za prijateľný interval spoľahlivosti označujeme taký interval hodnôt p , ktorý je v istom zmysle najmenšia množina s koeficientom spoľahlivosti (dosiahnutou pravdepodobnosťou pokrytia skutočnej hodnoty parametra), čo najbližšie k nominálnej hladine $1 - \alpha$. Preto hlavný dôraz pri vzájomnom porovnávaní intervalov kladieme na dosiahnutú pravdepodobnosť pokrytia, ďalej len *pravdepodobnosť pokrytia* (probability coverage). Ďalším prirodzeným

kritériom posudzovania intervalov je úspornosť (parsimony) meraná strednou dĺžkou intervalov. Ghosh (1979) [12] vyhodnocoval binomické intervaly spoľahlivosti použitím troch základných kritérií: pravdepodobnosť pokrytia, dĺžka intervalu a *Neyman shortness*, čo je pravdepodobnosť, ktorou interval pokrýva chybné hodnoty parametra $p' \neq p$, teda hovoríme o *pravdepodobnosti chybného pokrytia*. Vlastnosťou intervalov spoľahlivosti je tendencia pokrývať tieto chybné hodnoty s relatívne malými pravdepodobnosťami. Tak ako stojí chyba chyba prvého druhu voči chybe druhého druhu, stojí pravdepodobnosť pokrytia voči pravdepodobnosti chybného pokrytia. Neyman shortness sa zvyčajne neuvažuje pri hodnotení intervalov. Z časti môže za to fakt, že pri jej vyhodnocovaní dostávame neprehľadnú tabuľku pravdepodobností pre niekoľko vybraných hodnôt p' . *Koeficient strednej dĺžky EV* (expected volume), čo je vlastne priemer cez všetky stredné dĺžky (volume), zhŕňa v jednom čísle to, čo je nekompletné uvedené v niekoľkých tabuľkách a grafoch pre strednú dĺžku a pravdepodobnosť chybného pokrytia.

Preto navrhujeme numerické hodnotenie intervalov spoľahlivosti postaviť na základe *priemerného pokrytia* a priemernej strednej dĺžky. V niektorých prípadoch uvádzame aj *minimálne pokrytie*, čo je minimum z pravdepodobnosti pokrytia cez všetky možné hodnoty parametru p . Tento návrh ďalej aplikujeme na spomínané „odporúčané intervaly“ a štandardný interval.

V snahe o hlbšie študovanie vlastností intervalov spoľahlivosti navrhujeme uvažovať asymptotický rozvoj. Dohromady, Edgeworthov rozvoj pre pravdepodobnosť pokrytia, asymptotický rozvoj pre strednú dĺžku sú vhodnými nástrojmi na kompletne a vzájomne porovnateľné vyhodnotenie navrhovaných intervalov. Teoretická analýza prevedená v tejto časti má posilniť predchádzajúce zistenia a konečné závery.

4.1 Návrh na hodnotenie

Nech $X \sim \text{Bi}(n, p)$. Uvažujme CI ako interval spoľahlivosti s nominálnym koeficientom spoľahlivosti $1 - \alpha$, pre pevnú pravdepodobnosť úspechu p v postupnosti n nezávislých pokusov s počtom úspechov X . Potom interval spoľahlivosti je určený ľavým a pravým koncovým bodom intervalu, čo sú výrazy v premennej x , závislé na parametre n a predom určenej hodnote α , označme $L(x)$, respektíve $U(x)$. Koncové body intervalu sú náhodné veličiny, rovnako aj *dĺžka intervalu spoľahlivosti* je náhodná veličina. Platí nasledovná rovnosť

$$\text{dĺžka } CI(X) = U(X) - L(X) = \int_{p \in CI(X)} dp. \quad (4.1)$$

Ak je dĺžka jedno z kritérií hodnotenia intervalov, je prirodzené uvažovať *strednú dĺžku intervalu* (mean length). Strednú dĺžku intervalu definujeme ako

$$E\{\text{dĺžka } CI(X)\} = \sum_{x=0}^n (U(x) - L(x)) \binom{n}{x} p^x (1-p)^{n-x}. \quad (4.2)$$

Strednú dĺžku intervalu spoľahlivosti pre parameter p môžeme vyjadriť aj cez pravdepodobnosť pokrytia, a to nasledovne

$$\begin{aligned} E\{\text{dĺžka } CI(X)\} &= \int \int_{p \in CI} dp dP(x) \\ &= \int \int_{p \in CI} dP(x) dp = \int P\{p \in CI\} dp. \end{aligned} \quad (4.3)$$

V rovnosti (4.3) najprv jednoducho dosadíme z (4.1), miera P odpovedá distribučnej funkcii X , a nakoniec prehodíme poradie integrálov. Identita

$$\int P\{p \in CI\} dp = \int P\{p' \in CI\} dp' \quad (4.4)$$

hovorí o tom, že vynechaním jedinej hodnoty parametra p , z integrácie na pravej strane rovnosti (4.3), sa hodnota integrálu nezmení. Ak skutočná hodnota parametra je p , potom stredná dĺžka je rovná integrovanej pravdepodobnosti, že interval CI pokrýva chybnú hodnotu parametra p' . Aplikovaním identity (4.4) na binomické rozdelenie X , stredná dĺžka (4.2) môže byť vyjadrená ako

$$\int_0^1 \sum_{x: p' \in C(x)} \binom{n}{x} p^x (1-p)^{n-x} dp',$$

kde vidíme spojitosť medzi strednou dĺžkou a pravdepodobnosťou chybného pokrytia. Samozrejme pre počítanie je jednoduchšie vychádzať z (4.2).

Pravdepodobnosť pokrytia parametra p pre interval spoľahlivosti $CI(X)$ definujeme predpisom

$$C(p, n) \equiv P_p\{L(X, n) \leq p \leq U(X, n)\} = \sum_{x: p \in C(x)} \binom{n}{x} p^x (1-p)^{n-x}. \quad (4.5)$$

Vidíme, že pravdepodobnosť pokrytia sa mení v závislosti na hodnotách p a n . Štandardná podmienka pre interval spoľahlivosti je: najmenšia množina vyhovujúca požiadavke $\min_p P_p(p \in CI) \geq 1 - \alpha$, kde $\min_p P_p(p \in CI)$ nazývame *koeficient spoľahlivosti*. Štandardná podmienka vedie k výberu intervalu spoľahlivosti na základe neznámeho parametra. A pre každú pevnú

hodnotu parametra existuje najkratší interval. Pre pravdepodobnosť pokrytia uprednostňujeme alternatívnu podmienku: *priemerná pravdepodobnosť pokrytia* (average coverage) je najmenej rovná $1 - \alpha$. Zavedme priemernú pravdepodobnosť pokrytia predpisom

$$AC(CI) = \int_0^1 C(p, n) dp, \quad (4.6)$$

čo je intergrál z pravdepodobnosti pokrytia cez celý parametrický priestor $(0, 1)$. Uvažujeme rovnomerné rozdelenie p na $(0, 1)$, $p \sim R(0, 1)$. Rovnaký efekt by sme dosiahli zpriemerovaním $C(p, n)$ cez $p \in (0, 1)$ a uvažovali by sme iné ako rovnomerné pravdepodobnostné rozdelenie p . Jedno z často používaných je beta rozdelenie.

Pre určenie ako sa líši dosiahnutá pravdepodobnosť pokrytia od nominálneho koeficientu spoľahlivosti $1 - \alpha$ zavádzame odmocninu zo *strednej štvorcovej chyby* (root mean square error)

$$RMSE(CI) = \sqrt{\int_0^1 \{C(p, n) - (1 - \alpha)\}^2 dp}. \quad (4.7)$$

Dĺžku intervalu spoľahlivosti a pravdepodobnosť chybného pokrytia môžeme zhrnúť do jedného kritéria. Pre tento účel navrhujeme kritérium *koeficient strednej dĺžky EV* dané predpisom

$$EV(CI) = \int_0^1 E(\text{dĺžka } CI) dp. \quad (4.8)$$

Hoci $EV(CI)$ sumarizuje vlastnosti intervalov, stále závisí na pevných parametroch, ako sú n a α . Precíznejšie by bolo písať $EV(CI; \alpha, n)$ a rovnako $AC(CI; \alpha, n)$. Pri hodnotení intervalov spoľahlivosti musíme uvažovať alternatívny výber hodnôt n a α . Na rozdiel od p , ide však o známe parametry. Pre jednoduchosť sa ďalej obmedzíme na prípad $\alpha = 0,05$.

Na základe navrhovaných kritérií (4.6) a (4.8) za najlepší interval spoľahlivosti pokladáme interval CI s najmenším $EV(CI)$, ktorý spĺňa $AC(CI) \geq 1 - \alpha$. Hoci takýto interval spoľahlivosti sa pre rôzne hodnoty parametrov n a α líši, pre pevné parametry môžeme urobiť dôkladné a prehľadné tabuľky. Ako sa mení n , kritéria AC a EV sa menia tiež a potrebujú byť počítané pre každé n . Metódy konštruovania intervalov spoľahlivosti, ktoré pokladáme za dobré a spoľahlivé majú byť určitým kompromisom a majú viesť k intervalovým odhadom, ktorých uspokojivé hodnoty AC a EV pretrvávajú pre rôzne hodnoty n a α .

Pri posudzovaní intervalov sa zvykne uvažovať minimálne pokrytie. Zavedme *minimálnu pravdepodobnosť pokrytia* (minimum coverage) predpisom

$$MC(CI) = \min_p P_p(p \in CI).$$

Vieme, že pravdepodobnosť pokrytia intervalu spoľahlivosti pre p , ak $p = p_0$, je vlastne pravdepodobnosť odpovedajúceho oboru nezamietnutia nulovej hypotézy testu $p = p_0$, za platnosti nulovej hypotézy. Ďalej vieme, že s p rastúcim z 0 do 1, pravdepodobnosť $P(a \leq X \leq b)$ narastá do maxima a potom klesá. O tom sa ľahko presvedčíme deriváciou. Označme $f(p) = P(a \leq X \leq b)$, potom platí

$$f'(p) = \sum_{x=a}^b \binom{n}{x} p^{x-1} (1-p)^{n-x-1} (x - np).$$

Vidíme, že funkcia $f(p)$ je rastúca (kladná derivácia) pre $p < (a+b)/2n$ a klesajúca inak. Z toho máme, že minimum cez celý interval spoľahlivosti $CI(X)$ hodnôt p musí ležať niekde blízko koncových bodov interval. Potom $MC(CI(X))$ pre dané n musí byť niekde veľmi blízko $p = U(x, n)$, $p = L(x, n)$. Špeciálne, hľadáme minimum $MC(n)$ v jednom z bodov $p = L - \varepsilon$ alebo $p = U + \varepsilon$ pre $\varepsilon = 0$ alebo dostatočne malé.

Rovnako ako pri pravdepodobnosti pokrytia intervalov uvažujeme minimálne pokrytie, pre kritérium strednej dĺžky by sme uvažovali *maximálnu strednú dĺžku MV* (maximum volume)

$$MV(C(X)) = \max_p E\{\text{dĺžka } CI(X)\} = \max_{p'} \int P\{p' \in CI\} dp'.$$

Toto kritérium sa zdá byť vhodná náhrada za priemernú strednú dĺžku, ktoré nepožaduje predpoklad pravdepodobnostného rozdelenia p . Ale pri aplikovaní na binomický prípad, sa ukáže ako neúčinné. Pravdepodobnosť pokrytia intervalov spoľahlivosti v binomickom prípade je veľmi slabá pre p blízko 0 alebo 1 (viď napr. obrázok 4.1). Minimálne pokrytie (a tiež maximálne) sa vyskytuje v týchto krajných úsekoch. Z tohto a z druhej rovnosti vo vyššie uvedenej definícii kritéria MV vidíme, že toto kritérium dáva váhu stredným hodnotám p .

Všetky hodnoty uvedené v tabuľkách a grafoch sú počítané priamo z definícií pomocou funkcií **coverage** a **length**. Nie sú založené na simuláciách. Výpočty sú prevedené v programe R. Pre výpočty obidvoch kritérií využívam knihovňu QUADPACK zabudovanú v programe R. Pre AC integrujem vzorec (4.6) cez parametrický prostor $(0, 1)$ rozdelený do 1113 subintervalov, a pre EV integrujem formulu (4.8) cez $(0, 1)$ rozdelený do 7 subintervalov.


```

coverage <- function(n,p,alfa,metoda= c("Wald","Wilson",
                                         "Agresti", "Jefferys","LR"))
{
  indik <- 0:n
  lower <- 0:n
  upper <- 0:n
  pravd <- dbinom((0:n),n,p)
  for (i in 1:(n+1)){
    lower[i] <- metoda(i-1,n,p,alfa)[1]
    upper[i] <- metoda(i-1,n,p,alfa)[2]
    indik[i] <- ifelse(lower[i]<=p&upper[i]>=p,1,0)
  }
  vysledok <- pravd%*%indik
  return(vysledok)
}

length <- function(n,p,alfa,metoda=metoda= c("Wald","Wilson",
                                              "Agresti", "Jefferys","LR"))
{
  pravd <- dbinom((0:n),n,p)
  lg <- 0:n
  lower <- 0:n
  upper <- 0:n
  for (i in 1:(n+1)){
    lower[i] <- metoda(i-1,n,p,alfa)[1]
    upper[i] <- metoda(i-1,n,p,alfa)[2]
    lg[i] <- upper[i]-lower[i]
  }
  vysledok <- pravd%*%lg
  return(vysledok)
}

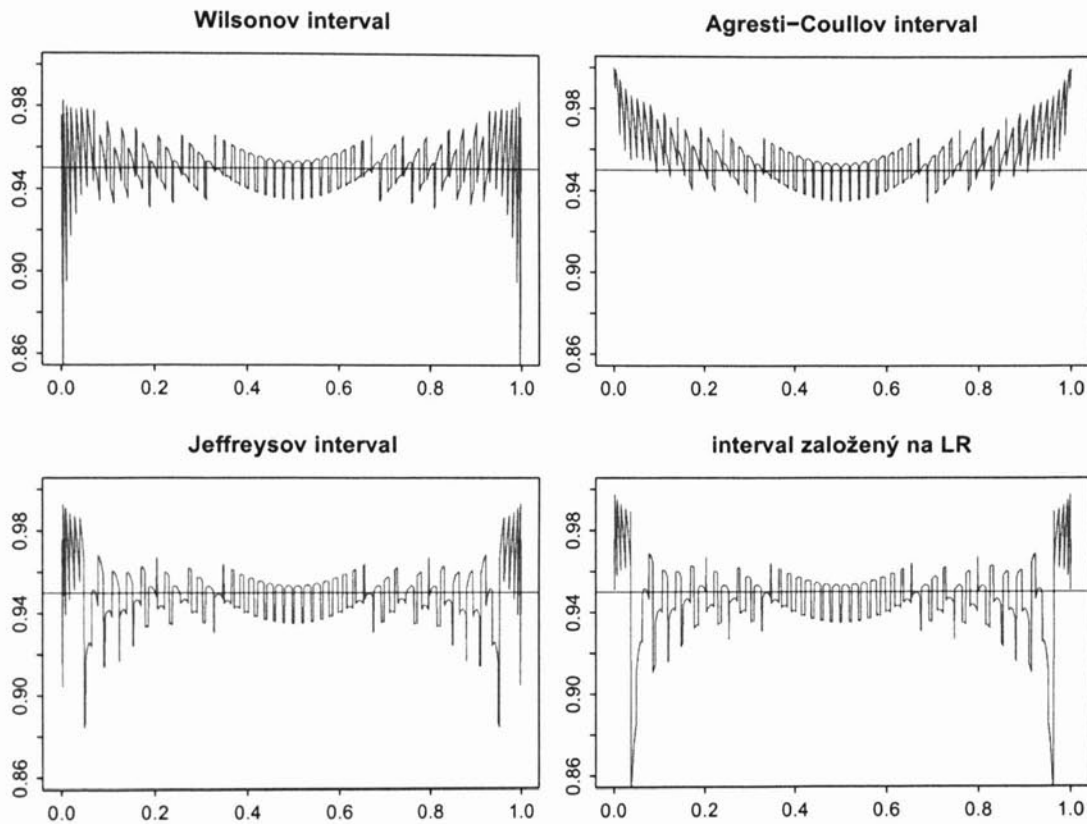
```

Pre popis jednotlivých metód ("Wald", "Wilson", "Agresti", "Jefferys") a ich zdrojového kódu v R viď kapitolu 3.

Vyhodnotenie

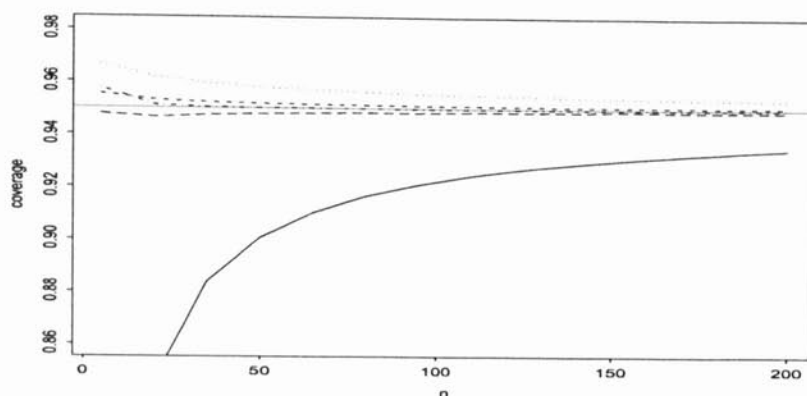
V tejto časti budeme porovnávať vlastnosti odporúčaných intervalov pomocou štandardných nástrojov (pravdepodobnosť pokrytia a stredná dĺžka) a pomocou vyššie navrhnutých kritérií *AC* a *EV*.

Obrázok 4.1 zobrazuje pravdepodobnosť pokrytia štyroch odporúčaných intervalov pre $n = 50$. Pri porovnaní s obrázkom 2.3 vidíme, že celkový trend

Obrázok 4.1: Pravdepodobnosť pokrytia pre $n = 50$ a $\alpha = 0,05$

kriviek na obrázku 4.1 je značne bližšie nominálnemu koeficientu spoľahlivosti ako je to u štandardného intervalu. Vidíme, že Wilsonov interval kolíše uspokojivo okolo $1 - \alpha$, s výnimkou úzkych okrajových oblastí p blízko 0 alebo 1. Na druhej strane Agresti-Coullov interval má tendenciu byť v týchto úsekoch značne väčší ako nominálny koeficient. V porovnaní s Wilsonovým intervalom má väčšiu pravdepodobnosť pokrytia. To nie je prekvapujúce, lebo ako sme sa už zmienili CI_{AC} vždy obsahuje CI_W ako vlastnú podmnožinu. Zostávajúce dva intervaly sú v priebehu pravdepodobnosti pokrytia porovnateľné, pričom priebeh pre Jeffreysov interval je zomknutejší k 0,95. Jeffreysov interval sa zdá byť vážnym kandidátom na najlepší interval spoľahlivosti v zmysle dosiahnutej pravdepodobnosti pokrytia, a to až na jeden hlboký pik blízko 0, a symetricky, druhý blízko 1.

Hoci grafy sú veľmi pôsobivé, nie sú najšťastnejšou voľbou na zobrazenie situácie pre rôzne hodnoty n . Tabuľky poskytujú ucelenejší prehľad, kde môžeme pozorovať celú škálu hodnôt n . Preto sú predmetom menšej chyby



Obrázok 4.2: Porovnanie priemernej pravdepodobnosti pokrytia pre Waldov interval (pevná čiara), Wilsonov (prerušovaná), Agresti-Coullov (bodkovaná), Jeffreysov (bodkočiarkovaná) a interval založený na LR (dlhé pomlčky). Nominálny koeficient spoľahlivosti je $\alpha = 0,05$

v záveroch, spôsobenej nevyčerpaním všetkých možností. V tabuľkách 4.1 a 4.2 sú uvedené hodnoty navrhnutých kritérií. Kritériá AC , EV sú vyčíslené v 7 rôznych hodnotách n v rozsahu od 5 do 1000. Ako prvé si všimneme, že priemerná pravdepodobnosť pokrytia pre Waldov interval má tendenciu byť výrazne nižšia. Napríklad $AC(CI_S) = 0,6406$ pre $n = 5$, v porovnaní s $AC(CI_{LR}) = 0,9477$. Automaticky to však neznamená, že Waldov interval je nevyhnutne užší. Konkrétne v tomto prípade máme $EV(CI_S) = 0,4600$ a $EV(CI_{LR}) = 0,4324$. Zaujímavým faktom je porovnanie AC a EV v tabuľkách 4.1 a 4.2. Zvyčajne ako $EV = EV(CI, n)$ narastá alebo klesá, potom $AC = AC(CI, n)$ narastá alebo klesá a naopak. V dôsledku toho by sme očakávali, že minimálna hodnota AC je pri vyhodnocovaní intervalov akousi alternatívou pre minimálnu hodnotu EV . Avšak z tabuliek vidíme, že tomu nie je vždy tak. Špeciálne vidíme, že $AC(CI_{LR}) < AC(CI_J)$ ale $EV(CI_{LR}) > EV(CI_J)$, pre väčšinu hodnôt n . Preto kritérium najmenej EV nemôžeme nahradiť kritériom najmenej AC . Pri rozhodovaní medzi menším AC a menším EV preferujeme splnenie požiadavky $AC \geq 1 - \alpha$. Obrázok 4.2 zobrazuje priemernú pravdepodobnosť pokrytia ako funkciu n . Demonštruje ostré rozdiely v pravdepodobnosti pokrytia medzi odporúčanými intervalmi. Na jednej strane krivka pre štandardný interval $AC(CI_S)$ leží výrazne pod hranicou 0,95, na druhej strane krivka pre Agresti-Coullov interval $AC(CI_{AC})$ je jemne nad. Agresti-Coullov interval je v zmysle priemerného pokrytia konzervatívny. Najbližšie k 0,95 sú hodnoty pre Jeffreysov interval. Stabilný priebeh pre celý rozsah n , aj keď jemne podhodno-

Tabuľka 4.1: Priemerná pravdepodobnosť pokrytia 95%-ných intervalov spoľahlivosti s odmocninou zo strednej štvorcovej chyby uvedenou v zátvorkách

n	5	15	50	100	200	1000
CI_S	0,6406 (0,3992)	0,8188 (0,2358)	0,9006 (0,1310)	0,9222 (0,0918)	0,9346 (0,0660)	0,9462 NA
CI_W	0,9553 (0,0293)	0,9534 (0,01676)	0,9518 (0,0100)	0,9511 (0,0096)	0,9507 (0,0065)	0,9504 (0,0029)
CI_{AC}	0,9665 (0,0284)	0,9629 (0,0197)	0,9580 (0,0145)	0,9555 (0,0118)	0,9536 (0,0096)	0,9512 (0,0032)
CI_J	0,9574 (0,0269)	0,9517 (0,0223)	0,9501 (0,0141)	0,9499 (0,0078)	0,9499 (0,0045)	0,9500 (0,0045)
CI_{LR}	0,9477 (0,0494)	0,9461 (0,0322)	0,9478 (0,0189)	0,9486 (0,0078)	0,9490 (0,0071)	0,9498 (0,0033)

Tabuľka 4.2: Koefficient strednej dĺžky „odporúčaných“ intervalov s nominálnym koefficientom spoľahlivosti 0,95

n	5	15	50	100	200	1000
CI_S	0,4600	0,3533	0,2113	0,1518	0,1082	0,0486
CI_W	0,5581	0,0000	0,3254	0,1523	0,1083	0,0486
CI_{AC}	0,5880	0,3839	0,2175	0,0154	0,1091	0,0487
CI_J	0,5606	0,3656	0,2117	0,1517	0,1080	0,0486
CI_{LR}	0,4324	0,3481	0,2100	0,1513	0,1079	0,0486

Tabuľka 4.3: Minimálna pravdepodobnosť pokrytia intervalov uvedená ako minimum cez $5 < n < 200$

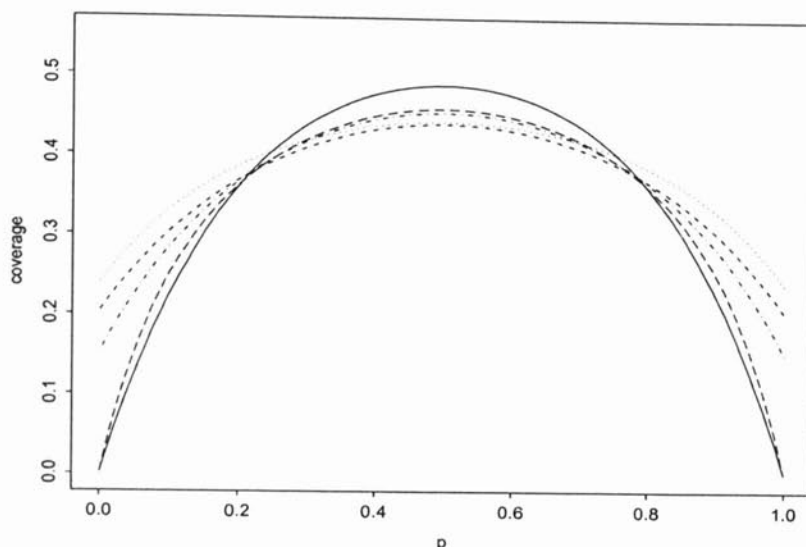
min	MC_{15}	MC_{30}	MC_{50}	MC_{100}	MC_{150}	n v min	$MC\%$
CI_S	0,0004	0,0013	0,7942	0,0060	0,7938	1	00,04
CI_W	0,9075	0,9216	0,9339	0,9311	0,9381	1	90,75
CI_{AC}	0,9167	0,9304	0,9340	0,9360	0,9392	2	91,67
CI_J	0,8731	0,9132	0,8869	0,8869	0,9328	4	87,31
CI_{LR}	0,8018	0,8232	0,8166	0,9086	0,9317	1	80,18

Minimum je bráné cez všetky hodnoty p . $MC(CI) = \min_p P_p(p \in CI)$, $MC_{15} = \min_{5 \leq n < 15} MC(CI)$, $MC_{30} = \min_{15 \leq n < 30} MC(CI)$, $MC_{50} = \min_{30 \leq n < 50} MC(CI)$, $MC_{100} = \min_{50 \leq n < 100} MC(CI)$, $MC_{150} = \min_{100 \leq n < 200} MC(CI)$

tený, má krivka intervalu založeného na pomere vierohodnosti, v kontraste s Wilsonovým intervalom, ktorého pokrytie je špeciálne pre malé hodnoty n nad nominálny koeficient spoľahlivosti. Pre číselné vyhodnotenie rozdielu medzi dosiahnutou pravdepodobnosťou pokrytia a minimálneho koeficientu spoľahlivosti počítame $RMSE(CI)$ definovanú v (4.7). Vidíme veľký rozdiel v pokrytí parametru p medzi štandardným intervalom a ostatnými intervalmi. Vzájomné rozdiely medzi CI_W , CI_{AC} , CI_J a CI_{LR} sú menšie, a z tabuľky 4.1 a grafu 4.2 môžeme vyčítať zoradenie intervalov v tomto poradí: Agresti-Coull interval, Wilsonov a Jeffreysov interval. V tabuľke 4.3 sú vyhodnotené intervaly spoľahlivosti v zmysle minimálnej pravdepodobnosti pokrytia. V porovnaní s tabuľkou 4.1 vidíme, že hoci priemerná pravdepodobnosť pokrytia pre CI_W , CI_{AC} a CI_J dosahuje nominálny koeficient spoľahlivosti, minimálna pravdepodobnosť pokrytia je veľmi nízka. Tento trend je zreteľný najmä pre Jeffreysov interval. Ako sme ukázali pri štandardnom intervale pravdepodobnosť pokrytia osciluje, a to v oboch premenných n i p . Tejto oscilácii sa nevyhneme ani pri ostatných metódach aplikovaných na binomické, a teda diskkrétne rozdelenie.

Preto minimálne pokrytie môže byť veľmi malé. Tohto minima sa vizuálne nadobúda v niektorom ostróm piku krivky pravdepodobnosti pokrytia, čo odpovedá jednému bodu, hoci pre ostatné hodnoty p pravdepodobnosť pokrytia môže byť oveľa vyššia.

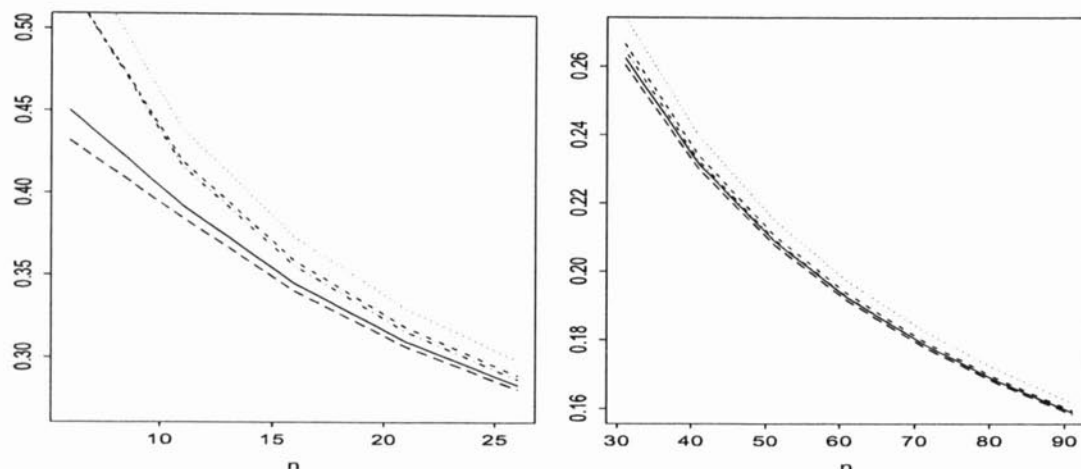
Požiadavok $P_p(p \in CI) = 1 - \alpha$ je ideálny, a zvyčajne splnený len pre



Obr. 4.3: Porovnanie strednej dĺžky pre Waldov interval (pevná čiara), Wilsonov (prerušovaná), Agresti-Coullov (bodkovaná), Jeffreysov (bodkočiarkovaná) a interval založený na LR (dlhé pomlčky) pre $\alpha = 0,05$ a $n = 15$

spojité rozdelenie. Štandardný požiadavok $P_p(p \in CI) \geq 1 - \alpha$ vedie k nadhodnotenému intervalovému odhadu, pre ktorý typická dosiahnutá pravdepodobnosť pokrytia je okolo 0,98 alebo 0,99. Preto sa preferuje interval spoľahlivosti, ktorý je úspornejší, ktorého priemerná pravdepodobnosť pokrytia dosahuje hladinu nominálneho koeficientu spoľahlivosti, ale pre niektoré hodnoty p môže byť menšia ako 0,95. V úvahách sa teda nahradzujeme požiadavok minimálna pravdepodobnosť pokrytia priemernou pravdepodobnosťou pokrytia.

Pre pevné n a p stredná dĺžka intervalového odhadu je užitočná miera kvality. Obrázok 4.3 zobrazuje strednú dĺžku piatich intervalov spoľahlivosti ako funkciu p pre $n = 15$. Pre tento prípad dostávame nasledujúce výsledky. Štandardný interval je najkratší špeciálne pre p blízko 0 alebo 1. Odráža to fakt, že ak $x = 0$ alebo $x = n$ tento interval je omedzený na 0, resp. 1. Obidva obrázky 4.3 a 4.4 podtrhávajú skutočnosť, že slabá pravdepodobnosť pokrytia štandardného intervalu nevyplýva z toho, že je príliš krátky. A Agresti-Coullov interval CI_{AC} je najdlhší pre väčšinu hodnôt p . Ďalej vidíme, že Agresti-Coullov interval je špeciálne pre malé n o dosť väčší, čo má za následok jemne vyššiu pravdepodobnosť pokrytia ako sme si mohli všimnúť v tabuľke 4.1. Z grafu 4.4 sa dá pekne vyčítať zoradenie intervalov v ich priemernej strednej dĺžke. A tak vidíme, že interval založený na vierohod-



Obrázok 4.4: Koeficient strednej dĺžky intervalov spoľahlivosti pre $\alpha = 0,05$ a n v rozsahu 5 až 30 a 31 až 100. Vid' popis obrázka 4.3

nostnom pomere je najúspornejší. S rastúcim n sa rozdiely medzi strednou dĺžkou intervalov strácajú.

Na základe hodnotenia postaveného na kritériách (4.6), (4.8) z alternatívnych intervaloch za najlepší interval spoľahlivosti pokladáme Jeffreysov interval. Aj keď z obrázka 4.4 vidíme, že v zmysle priemernej strednej dĺžky nie je najúspornejší, pri hodnotení intervalov uprednostňujeme kritérium AC . Dobré vlastnosti Jeffreysovho intervalu pretrvávajú aj pre malé hodnoty n . Pre veľké hodnoty n ($n \geq 100$) sa rozdiely medzi intervalmi takmer strácajú a intervaly spoľahlivosti CI_W , CI_{AC} , CI_J , CI_{LR} sú vzájomné porovnateľné.

4.2 Edgeworthov rozvoj

V celej tejto časti budem naďalej porovnávať štandardný interval a štyri odporúčané intervaly, a to v ich *pravdepodobnosti pokrytia* definovanej v (4.5). Pravdepodobnosťou pokrytia intervalov sa budeme zaoberať podrobnejšie z tej analytickej stránky prostredníctvom Edgeworthovho rozvoja. Týmto sa majú posilniť výsledky numerickej analýzy intervalov prevedenej v predchádzajúcej časti.

Edgeworthov rozvoj je populárny nástroj na študovanie komplikovaných vlastností v štatistike. Edgeworthov rozvoj je prehľadne spracovaný v Bhattacharya a Rao (1976) [5] a Hall (1992) [14]. Osobitne Edgeworthov rozvoj je užitočný nástroj na porozumenie oscilačného javu v priebehu pravdepo-

dobnosti pokrytia. Tu budeme odvodením príslušného Edgeworthovo rozvoja študovať pravdepodobnosť pokrytia alternatívnych intervalov. Malo by sa poznamenať, že môže byť použitý aj iný typ asymptotického rozvoja, ale ako uvidíme neskôr, Edgeworthov rozvoj sa zdá byť najvhodnejší, pretože výraznejšie zachycuje oscilačný jav. Odvodenie príslušného Edgeworthovho rozvoja je o niečo viac technické, nevýhodou je zdĺhavá a pre značenie neprijemná algebra. Dôkazy a odvodenia sú uvedené v dodatku.

O Edgeworthovom rozvoji pojednáva nasledovná veta. Uvažujme náhodnú veličinu X_1 s charakteristickou funkciou ψ_1 , so strednou hodnotou $E(X_1) = 0$, rozptylom $E(X_1^2) = 1$, a $E(|X_1|^{r+2}) < \infty$, pre nejaké celé $r \geq 0$. Označme κ_j kumulanty X_1 a $K_1(t)$ kumulatívnu vytvárajúcu funkciu, pričom $\kappa_0 = 0 = \kappa_1, \kappa_2 = 1$. Taylorovým rozvojom kumulatívnej vytvárajúcej funkciu $K_1(t) := \ln(\psi_1(t))$ dostávame

$$\begin{aligned} K_1(t) &= \sum_{j=0}^{r+2} \frac{\kappa_j (it)^j}{j!} + o(t^{r+2}) \\ &\approx -\frac{t^2}{2} + \frac{\kappa_3 (it)^3}{3!} + \dots + \frac{\kappa_{r+2} (it)^{r+2}}{(r+2)!} := -\frac{t^2}{2} + \beta(t), \end{aligned}$$

ako $t \rightarrow \infty$. Teraz nech $X_n, n \in \mathbb{N}$, sú nezávislé rovnako rozdelené náhodné veličiny s rozdelením ako X_1 . Položme

$$S_n = X_1 + X_2 + \dots + X_n \quad \text{a} \quad S_n^* = S_n / \sqrt{n}. \quad (4.9)$$

Potom S_n^* má kumulatívnu vytvárajúcu funkciu

$$K_n(t) = nK_1(t/\sqrt{n}) \approx -t^2/2 + \beta(t/\sqrt{n}) := \widehat{K_n(t)}$$

a charakteristickú funkciu

$$\psi_n(t) = (\psi_1(t/\sqrt{n}))^n \approx e^{\widehat{K_n(t)}} := \widehat{\psi_n(t)}.$$

Ak $\psi_n(t)$ je integrovateľná, potom hustota S_n^* je daná inverznou formulou

$$\begin{aligned} f_n(x) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \psi_n(t) dt \\ &\approx \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \widehat{\psi_n(t)} dt := \widehat{f_n(x)}. \end{aligned} \quad (4.10)$$

Veta 4.1 (Edgeworthov rozvoj) *Nech X_1, X_2, \dots sú nezávislé rovnako rozdelené náhodné veličiny s $E(X_1) = 0$, $E(X_1^2) = 1$, a $E(|X_1|^{r+2}) < \infty$, pre nejaké celé $r \geq 0$. Predpokladajme, že pre nejaké $\nu \in \mathbb{N}$ platí*

$$\int_{-\infty}^{\infty} |\psi_1(t)|^\nu dt < \infty.$$

Definujme $\widehat{f_n(x)}$ ako v (4.10). Položme $S_n^* = S_n/\sqrt{n}$ ako v (4.9). Potom pre $n \geq \nu$, S_n^* má spojitú hustotu $f_n(x)$ takú, že pre $n \rightarrow \infty$ platí

$$\sup\{(1 + |x|^{2+r})|f_n(x) - \widehat{f_n(x)}| : x \in \mathbb{R}\} = o(n^{-r/2}). \quad (4.11)$$

Dôkaz: viď Wichura (2001) [24] Veta 1. ■

Funkcia $\widehat{f_n(x)}$ z vety 4.1 sa označuje ako Edgeworthov rozvoj druhého rádu hustoty S_n^* . Z toho tiež vyplýva (viď Wichura (2001) [24]) *Edgeworthov rozvoj r -tého rádu* distribučnej funkcia S_n^* daný predpisom

$$P\{S_n^* \leq x\} \approx \Phi(x) - \phi(x) \left[\sum_{j=1}^r p_j(x) \frac{1}{n^{j/2}} \right], \quad (4.12)$$

kde $\phi(x)$, resp. $\Phi(x)$ sú po rade hustota a distribučná funkcia štandardného normálneho rozdelenia definované v (2.1) a (2.2) a $p_j(x)$ sú polynómy definované v [24]. Tu uvádzame explicitný tvar len pre polynómy p_1 a p_2 , ktoré vystupujú v Edgeworthovom rozvoji 2. rádu

$$\begin{aligned} p_1(x) &= \frac{\kappa_3}{3!} H_2(x), \\ p_2(x) &= \frac{\kappa_4}{4!} H_3(x) + \frac{10\kappa_3^2}{6!} H_5(x), \end{aligned}$$

kde H_k sú takzvané *Hermitovské polynómy* dané predpisom

$$\begin{aligned} H_0(x) &= 1, \\ H_1(x) &= x, \\ H_k(x) &= xH_{k-1}(x) - (k-1)H_{k-2}(x), \quad \text{pre } k \geq 2. \end{aligned}$$

Z (4.11) vyplýva, že Edgeworthov rozvoj r -tého rádu aproximuje príslušnú distribučnú funkciu s presnosťou rádu $n^{-r/2}$. Poznamenajme, že Edgeworthov rozvoj 0-tého rádu všeobecnej distribučnej funkcie je štandardná normálna distribučná funkcia. Na rozvoj vyšších rádoov sa môžeme pozerat ako na určitú vylepšenú normálnu aproximáciu. Pre každé j sú p_j polynómy v premennej x stupňa $3j$ s koeficientami závisiacimi na kumulantoch κ_3 až κ_{j+2} . Kumulanty môžeme vyjadriť pomocou centrálnych momentov. Platí $\kappa_3 = \mu_3$ a $\kappa_4 = \mu_4 - 3\mu_2^2$, kde μ_r je r -tý centrálny moment. A tak κ_3 a κ_4 po rade odpovedajú šikmosti a špicatosti rozdelenia, z čoho vyplýva, že člen rozvoja rádu $n^{-1/2}$ v prvom rade reprezentuje chybu šikmosti, zatiaľ čo člen rozvoja rádu n^{-1} v prvom rade odráža špicatosť a až v druhom šikmosť. Definujeme

$$v(x) = x - [x],$$

kde $\lfloor x \rfloor$ je celá časť čísla x . Teda $v(x)$ je desatinná časť x . Funkcia $v(\cdot)$ je tak periodická periódou 1. Zavedme funkciu g predpisom

$$g(p, z) = g(p, z, n) = v(np + z\sqrt{npq}), \quad (4.13)$$

čím tiež zdôrazníme závislosť $g(p, z)$ na n .

Nech $X \sim \text{Bi}(n, p)$ a $\hat{p} = X/n$. Označme $S_n = n^{1/2}(\hat{p} - p)/(\sqrt{pq})$ a $F_n(z) = P(S_n \leq z)$. Veta 23.1 v Bhattacharya a Rao (1976) [5] vedie priamo k Edgeworthovmu rozvoju 2. rádu distribučnej funkcie F_n , ktorý je

$$\begin{aligned} F_n(z) = & \Phi(z) + \left(\frac{1}{6} - \frac{p}{3}\right) (1 - z^2)\phi(z) \frac{1}{\sqrt{pq}} n^{-1/2} \\ & + \left(\frac{1}{2} - g(p, z)\right) \phi(z) \frac{1}{\sqrt{pq}} n^{-1/2} \\ & + \left\{ \left(\frac{pq}{18} - \frac{1}{72}\right) \kappa^5 + \left(\frac{7}{72} - \frac{11pq}{36}\right) \kappa^3 + \left(\frac{pq}{12} - \frac{1}{12}\right) \kappa \right\} \phi(z) \frac{1}{pq} n^{-1} \\ & + \left\{ \left(\frac{1}{6} - \frac{p}{3}\right) (z^3 - 3) \left(\frac{1}{2} - g(p, z)\right) \right. \\ & \left. - \left(\frac{1}{2} g^2(p, z) - \frac{1}{2} g(p, z) + \frac{1}{6}\right) \right\} z \phi(z) \frac{1}{pq} n^{-1} + O(n^{-3/2}). \end{aligned} \quad (4.14)$$

Nech CI je pre ľubovoľný interval spoľahlivosti pre parameter p . Uvažujme konštruovanie intervalu spoľahlivosti CI s koeficientom spoľahlivosti $1 - \alpha$. Nech l a u sú výrazy v premennej p (a samozrejme n a κ) vyhovujúce nerovnosti

$$\{p \in CI\} = \left\{ l \leq \frac{n^{1/2}(\hat{p} - p)}{(pq)^{1/2}} \leq u \right\}. \quad (4.15)$$

Potom Edgeworthov rozvoj 2. rádu pravdepodobnosti pokrytia pre intervalu spoľahlivosti CI definujeme predpisom

$$P = P_p(p \in CI) = F_n(u) - F_n(l), \quad (4.16)$$

a $F_n(u)$ a $F_n(l)$ plynú z (4.14). V ďalšom budeme pravdepodobnosť pokrytia označovať skrátene ako $P = P_p(p \in CI)$. Nech $g(p, \cdot)$ je definovaná v (4.13). Kvôli zjednodušeniu zápisu zavedme nasledovné funkcie.

Značenie:

$$\begin{aligned} w(\kappa) &= \left(\frac{1}{9} - \frac{1}{36pq}\right) \kappa^5 + \left(\frac{7}{36pq} - \frac{11}{18}\right) \kappa^3 + \left(\frac{1}{6} - \frac{1}{6pq}\right) \kappa, \\ Q_1(l, u) &= 1 - g(p, l) - g(p, u), \\ Q_2(l, u) &= \frac{1}{2} \left(-g^2(p, l) - g^2(p, u) + g(p, l) - g(p, u) - \frac{1}{3} \right). \end{aligned}$$

Poznamenajme, že funkcia Q_2 nadobúda hodnoty medzi $-1/6$ a $1/12$ a Q_1 nadobúda hodnoty medzi -1 a 1 . Použitím tohto značenia rozpišeme (4.16) podľa (4.14) a dostávame

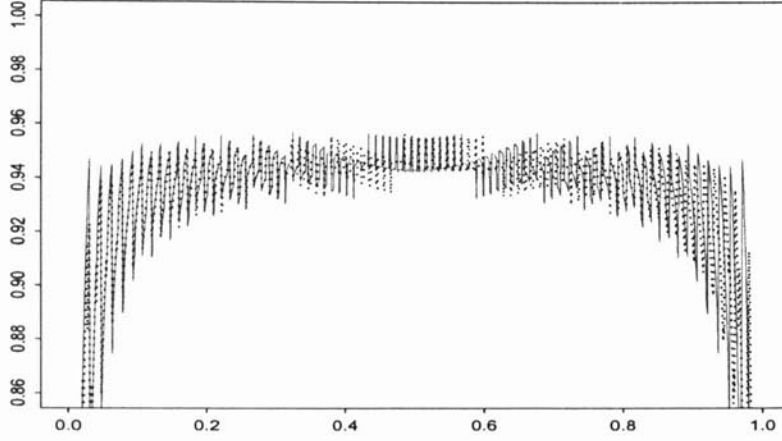
$$\begin{aligned} P = & 1 - \alpha + [g(p, l) - g(p, u)]\phi(\kappa)(npq)^{-1/2} + w(\kappa)\phi(\kappa)n^{-1} \\ & + \left\{ (1 - 2p) \left(\frac{\kappa^2}{6} - \frac{1}{2} \right) \cdot Q_1(l, u) + Q_2(l, u) \right\} \kappa\phi(\kappa)(npq)^{-1} \\ & + O(n^{-3/2}). \end{aligned}$$

Edgeworthov rozvoj 2. rádu pravdepodobnosti pokrytia môžeme symbolicky vyjadriť ako

$$P = 1 - \alpha + Osc_1 \cdot n^{-1/2} + S \cdot n^{-1} + Osc_2 \cdot n^{-1} + O(n^{-3/2}), \quad (4.17)$$

kde člen rozvoja $S \cdot n^{-1}$ (smooth term), zachytáva systematické vychýlenie v priebehu pravdepodobnosti pokrytia, ktoré sme mohli pozorovať na obrázku (2.5). Členy $Osc_1 \cdot n^{-1/2}$ a $Osc_2 \cdot n^{-1}$ (oscillatory term) odpovedajú chybe z nespojitosti binomického rozdelenia a sú pôvodcom oscilačného javu v pravdepodobnosti pokrytia. Pri konštruovaní obojstranných intervalov spoľahlivosti sa chyba plynúca z nespojitosti stáva dominantnou. Ukázalo sa, že Edgeworthov rozvoj je užitočný nástroj na hľadanie zdroja oscilácie. Hoci oscilačný člen vieme určiť, ak je p známe, z tvaru funkcie g a výrazov l a u , pre jednotlivé metódy uvedených v dodatku, vidíme, že ak je p neznáme, je nepredvídateľný. A tak pre p neznáme, nemáme žiadnu garanciu, že pravdepodobnosť pokrytia dosiahne predom stanovený nominálny koeficient spoľahlivosti. Ak by sme uvažovali Edgeworthov rozvoj pravdepodobnosti pokrytia 1. rádu (rozvoj, ktorý aproximuje distribučnú funkciu s presnosťou rádu $n^{-1/2}$), potom neoscilujúci člen $S \cdot n^{-1}$ zanedbávame. Avšak pre malé, dokonca aj stredné hodnoty n , tento člen nemôžeme zanedbať. Preto uvažujeme Edgeworthov rozvoj 2. rádu. Edgeworthov rozvoj 2. rádu aproximuje skutočnú pravdepodobnosť pokrytia intervalu spoľahlivosti s chybou $O(n^{-3/2})$. Táto aproximácia je dostatočne presná ako môžeme vidieť na obrázku 4.5. V ďalšom budeme uvažovať už len rozvoj 2. rádu.

Teraz uvidíme Edgeworthov rozvoj pravdepodobnosti pokrytia pre jednotlivé obojstranné intervaly spoľahlivosti z kapitoly 3. Používajme zavedené značenie $\phi(x)$ a $\Phi(x)$ pre hustotu, resp. distribučnú funkciu štandardného normálneho rozdelenia a κ pre $100(1 - \alpha)\%$ -ný kvantil normálneho rozdelenia. Nech $0 < \alpha < 1$ a predpokladajme, že p je vnútorný bod parametrického priestoru $0 < p < 1$. Kvantily l a u sú popísané v (4.15) a pre jednotlivé metódy formálne definované v dodatku.



Obrázok 4.5: Porovnanie medzi pravdepodobnosťou pokrytia (pevná čiara) a Edgeworthovým rozvojom 2. rádu pravdepodobnosti pokrytia (bodkovaná čiara) pre štandardný interval s $n = 100$

Veta 4.2 *Nech kvantily l_S a u_S sú definované ako (A.7) v dodatku. Predpokladajme, že $np + l_S(npq)^{-1/2}$ nie je celé číslo. Potom pravdepodobnosť pokrytia pre štandardný interval (2.3) je daná vzorcom*

$$\begin{aligned}
 P_S = P_p(p \in CI_S) &= 1 - \alpha + [g(p, l_S) - g(p, u_S)]\phi(\kappa)\frac{1}{\sqrt{pq}}n^{-1/2} \\
 &+ \left\{ -\frac{(1-2p)^2}{12pq}\kappa^5 - \frac{1}{4pq}\kappa^3 + w(\kappa) \right\} \phi(\kappa)(n)^{-1} \\
 &+ \left\{ -(1-2p) \left(\frac{\kappa^2}{3} + \frac{1}{2} \right) \cdot Q_1(l_S, u_S) + Q_2(l_S, u_S) \right\} \kappa\phi(\kappa)\frac{1}{pq}n^{-1} \\
 &+ O(n^{-3/2}).
 \end{aligned} \tag{4.18}$$

Člen $\left\{ -\frac{(1-2p)^2}{12pq}\kappa^5 - \frac{1}{4pq}\kappa^3 + w(k) \right\} \phi(\kappa)(n)^{-1}$ je v (4.18) kľúčovým výrazom.

Je neoscilujúci a je zdrojom systematickej negatívnej výchylky. Edgeworthov rozvoj pre Wilsonov interval je jednoduchší.

Veta 4.3 *Predpokladajme, že $np - \kappa(npq)^{-1/2}$ nie je celé číslo. Potom pravdepodobnosť pokrytia pre Wilsonov interval spoľahlivosti definovaný v (3.1) je daná vzorcom*

$$\begin{aligned}
P_W = \mathbf{P}_p(p \in CI_W) &= 1 - \alpha + [g(p, -\kappa) - g(p, \kappa)]\phi(\kappa)\frac{1}{\sqrt{pq}}n^{-1/2} \\
&+ w(\kappa)\phi(\kappa)n^{-1} + \left\{ (1 - 2p) \left(\frac{\kappa^2}{6} - \frac{1}{2} \right) \cdot Q_1(-\kappa, \kappa) \right. \\
&\left. + Q_2(-\kappa, \kappa) \right\} \kappa\phi(\kappa)\frac{1}{pq}n^{-1} + O(n^{-3/2}). \tag{4.19}
\end{aligned}$$

Podobne môžeme odvodiť Edgeworthov rozvoj 2. rádu pre Agresti-Coullov interval.

Veta 4.4 *Nech l_{AC} a u_{AC} sú definované ako v (A.8) v appendixe. Predpokladajme, že $np + l_{AC}(npq)^{-1/2}$ nie je celé číslo. Potom pravdepodobnosť pokrytia pre Agresti-Coullov interval spoľahlivosti zavedený v (3.5) je daná vzorcom*

$$\begin{aligned}
P_{AC} = \mathbf{P}_p(p \in CI_{AC}) &= 1 - \alpha + [g(p, l_{AC}) - g(p, u_{AC})]\phi(\kappa)\frac{1}{\sqrt{pq}}n^{-1/2} \\
&+ \left\{ \left(\frac{1}{4pq} - 1 \right) \kappa^3 + w(\kappa) \right\} \phi(\kappa)n^{-1} \\
&+ \left\{ (1 - 2p) \left(\frac{\kappa^2}{6} - \frac{1}{2} \right) \cdot Q_1(l_{AC}, u_{AC}) + Q_2(l_{AC}, u_{AC}) \right\} \\
&\kappa\phi(\kappa)\frac{1}{pq}n^{-1} + O(n^{-3/2}). \tag{4.20}
\end{aligned}$$

Nasledujúce vety nám dávajú Edgeworthov rozvoj pre pravdepodobnosť pokrytia Jeffreysovho intervalu a intervalu založeného na vierohodnostnom pomere. Aj tu môžeme príslušný Edgeworthov rozvoj odvodiť, ale problém je o niečo komplexnejší. Vyplýva to z faktu, že nemáme explicitné vyjadrenie pre koncové body týchto intervalov. Preto rozvoj spočíva v dvoch fázach. V prvej fáze rozvineme do asymptotickej rady koncové body intervalov samotných a až potom prevedieme vlastný Edgeworthov rozvoj 2. rádu. Výsledný rozvoj je uvedený v nasledujúcich vetách.

Veta 4.5 *Nech l_J a u_J sú definované ako v (A.14) v dodatku. Predpokladajme, že $np + l_J(npq)^{-1/2}$ nie je celé číslo. Potom pravdepodobnosť pokrytia pre Jeffreysov interval spoľahlivosti (3.6) je daná vzorcom*

$$\begin{aligned}
P_J = \mathbf{P}_p(p \in CI_J) &= 1 - \alpha + [g(p, l_J) - g(p, u_J)]\phi(\kappa)\frac{1}{\sqrt{pq}}n^{-1/2} \\
&- \frac{1}{12pq}\kappa\phi(\kappa)n^{-1} + \left\{ \frac{2p-1}{3} \cdot Q_1(l_J, u_J) + Q_2(l_J, u_J) \right\} \\
&\kappa\phi(\kappa)\frac{1}{pq}n^{-1} + O(n^{-3/2}).
\end{aligned} \tag{4.21}$$

Veta 4.6 *Nech l_{LR} a u_{LR} sú definované ako v (A.12) v dodatku. Predpokladajme, že $np + l_{LR}(npq)^{-1/2}$ nie je celé číslo. Potom pravdepodobnosť pokrytia pre interval spoľahlivosti založený na vierohodnostnom pomere (3.8) splňuje*

$$\begin{aligned}
P_{LR} = \mathbf{P}_p(p \in CI_{LR}) &= 1 - \alpha + \{g(p, l_{LR}) - g(p, u_{LR})\}\phi(\kappa)\frac{1}{\sqrt{pq}}n^{-1/2} \\
&+ \left(\frac{1}{6} - \frac{1}{6pq} \right) \kappa\phi(\kappa)n^{-1} + \left\{ \left(p - \frac{1}{2} \right) \cdot Q_1(l_{LR}, u_{LR}) \right. \\
&\left. + Q_2(l_{LR}, u_{LR}) \right\} \kappa\phi(\kappa)\frac{1}{pq}n^{-1} + O(n^{-3/2}).
\end{aligned} \tag{4.22}$$

Teraz použijeme Edgeworthov rozvoj na porovnanie vlastností pokrytia štandardného intervalu CI_S a štyroch alternatívnych intervalov, Wilsonovho intervalu CI_W , Agresti-Coullého intervalu CI_{AC} , Jeffreysovho intervalu CI_J a intervalu založeného na vierohodnostnom pomere CI_{LR} . Edgeworthov rozvoj rozkladá pravdepodobnosť pokrytia na päť častí (4.17). Intervaly porovnávam a vyhodnocujem na základe neoscilujúceho člena $S.n^{-1}$, ktorý je akousi mierou systematického vychýlenia v pravdepodobnosti pokrytia. Môžeme ho považovať za akúsi hladkú aproximáciu k pravdepodobnosti pokrytia. Koeficienty u člena n^{-1} určujú smer a veľkosť systematického vychýlenia v pravdepodobnosti pokrytia pre špecifický interval. Označme členy $S.n^{-1}$ v Edgeworthovom rozvoji pravdepodobnosti pokrytia intervalov CI_S , CI_W , CI_{AC} , CI_J a CI_{LR} po rade ako S_S , S_W , S_{AC} , S_J a S_{LR} . Ich tvar vyplýva priamo z (4.18), (4.19), (4.20), (4.21) a (4.22). Ich vzájomným porovnaním môžeme urobiť zaujímavé závery. Určujú, či pravdepodobnosť pokrytia konkrétneho intervalu má menšie vychýlenie ako interval, s ktorým ho porovnávame. Najpr ukážeme ako môžeme použiť koeficienty u člena n^{-1} na vysvetlenie ne-

dostatkov v pokrytí štandardnej metódy. Platí

$$S_W - S_S = \left\{ \frac{(1-2p)^2}{12pq} \kappa^5 + \frac{1}{4pq} \kappa^3 \right\} \phi(\kappa) n^{-1}, \quad (4.23)$$

$$S_{AC} - S_S = \left\{ \frac{(1-2p)^2}{12pq} \kappa^5 + \left(\frac{1}{2pq} - 1 \right) \kappa^3 \right\} \phi(\kappa) n^{-1}, \quad (4.24)$$

$$S_{LR} - S_S = \left\{ \frac{(1-2p)^2}{9pq} \kappa^5 + \left(\frac{1}{18pq} + \frac{11}{18} \right) \kappa^3 \right\} \phi(\kappa) n^{-1}, \quad (4.25)$$

$$S_J - S_S = \left\{ \frac{(1-2p)^2}{9pq} \kappa^5 + \left(\frac{1}{18pq} + \frac{11}{18} \right) \kappa^3 + \left(\frac{1}{12pq} - \frac{1}{6} \right) \kappa \right\} \phi(\kappa) n^{-1}. \quad (4.26)$$

Podstatná je skutočnosť, že v (4.23) a (4.24) sú koeficienty u člena n^{-1} kladné pre všetky p a všetky κ . Na základe čoho môžeme očakávať, že Wilsonov interval a vo väčšej miere Agresti-Coullov interval vylepšujú štandardný interval, a to z hľadiska systematického vychýlenia v pravdepodobnosti pokrytia. Výrazy (4.25) a (4.26) sú za predpokladu, že $\kappa > \sqrt{11/8} = 1,17$ takisto kladné, čo je väčšinou v praxi splnené. A tak aj tieto dva intervaly rovnako vylepšujú štandardnú metódu.

Teraz obdobným spôsobom porovnáme alternatívne intervaly navzájom. Priamo z (4.18)-(4.22) máme

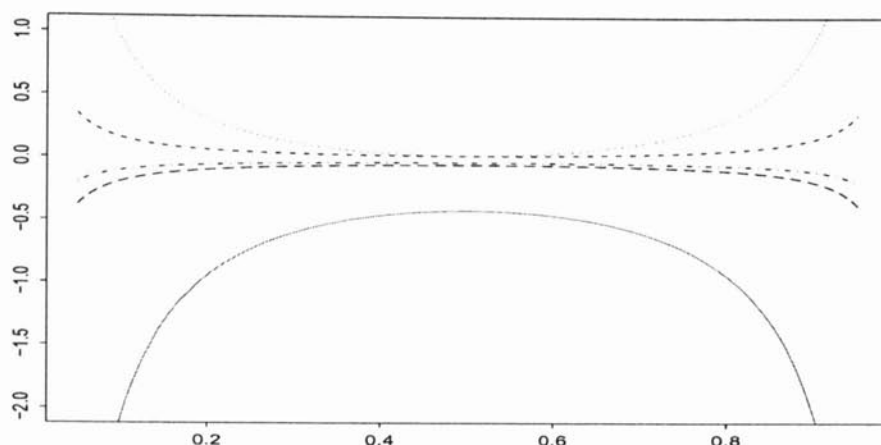
$$S_{AC} - S_W = \left(\frac{1}{4pq} - 1 \right) \kappa^3 \phi(\kappa) n^{-1}, \quad (4.27)$$

$$S_{AC} - S_{LR} = \left\{ -\frac{(1-2p)^2}{36pq} \kappa^5 + \left(\frac{4}{9pq} - \frac{29}{18} \right) \kappa^3 \right\} \phi(\kappa) n^{-1}, \quad (4.28)$$

$$S_{AC} - S_J = \left\{ -\frac{(1-2p)^2}{36pq} \kappa^5 + \left(\frac{4}{9pq} - \frac{29}{18} \right) \kappa^3 + \left(\frac{1}{6} - \frac{1}{12pq} \right) \kappa \right\} \phi(\kappa) n^{-1}. \quad (4.29)$$

V (4.27) je koeficient u n^{-1} kladný pre všetky p a všetky κ . Rovnako v (4.28) a (4.29) sú koeficienty u člena n^{-1} kladné pokiaľ $\kappa \leq 3,95$. Namiesto uvádzania takýchto explicitných výrazov, môžeme lepšie porozumieť rozdielom v jednotlivých metódach na základe obrázka 4.6.

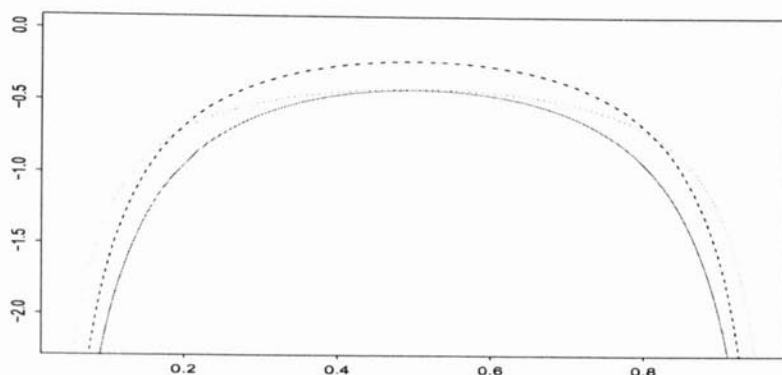
Obrázok 4.6 zobrazuje neoscilujúci člen rádu n^{-1} rozvoja pravdepodobnosti pokrytia pre každý interval. Hodnoty neoscilujúceho výrazu sú nanášané na osu y ako funkcie p pri danom pevnom $\alpha = 0,05$. Je zjavné, že štandardný interval je postihnutý nepretržitým systematickým negatívnym vychýlením v pravdepodobnosti pokrytia. Ďalej pozorujeme pozitívne vychýlenie pre Agresti-Coullov interval, viditeľné najmä blízko 0 alebo 1. Avšak



Obrázok 4.6: Porovnanie neoscilujúcich členov rádu n^{-1} v Edgeworthovom rozvoji pravdepodobnosti pokrytia pre Waldov interval (pevná čiara), Wilsonov (prerušovaná), Agresti-Coullov (bodkovaná), Jeffreysov (bodkočiarkovaná) a interval založený na LR (dlhé pomlčky) s $\alpha = 0,05$

najzaujímavejším faktom, ktorý odhaľuje obrázok 4.6, je takmer nulové systematické vychýlenie u intervalu založenom na pomere vierohodnosti a rovnako ako u Jeffreysovho intervalu. Edgeworthovo rozvoj a osobitne obrázok 4.6 ukazujú, že tieto dva intervaly majú takmer nulové vychýlenie v pravdepodobnosti pokrytia. Tiež vidíme, že vychýlenie Wilsonovho intervalu je jemne pozitívne a čo do veľkosti porovnateľné s negatívnym vychýlením u intervalu založenom na vierohodnostnom pomere.

Tieto porovnania jasne demonštrujú, že najväčšie pokrytie má Agresti-Coullov interval. Avšak najlepšie pokrytie parametra p , v zmysle najlepšieho priblíženia sa k danému nominálnemu koeficientu spoľahlivosti po zanedbaní oscilácie má Jeffreysov interval. V celom pojednávaní o alternatívnych intervaloch sa obmedzujeme len na 95% prípad, ale tu sa zamyslíme ako voľba nominálneho koeficientu spoľahlivosti ovplyvňuje kvalitu pokrytia parametra p . Uvažujme štandardný interval. Koeficient u člena n^{-1} je pre $\alpha = 0,05$ rovnomerne negatívnejší ako pre $\alpha = 0,01$ a $\alpha = 0,2$, čo poukazuje na to, že celý nominálny 95% interval spoľahlivosti je vo všeobecnosti viac vychýlený ako nominálny 99% alebo 80% interval. Tieto rozdiely môžeme pozorovať na obrázku 4.7. Aspoň vo forme poznámky uvediem ďalšiu skutočnosť vyplývajúcu z Edgeworthovho rozvoja. Teraz vezmime v úvahu pre zmenu oscilačné členy rozvoja. Oscilačný člen rozvoja pre štandardný interval je v dôsledku prítomnosti výrazu $\phi(\kappa)$ väčší pre $\kappa = 1,96$ ako pre $\kappa = 2,575$. To vysvetľuje



Obrázok 4.7: Porovnanie neoscilujúcich členov rádu n^{-1} v Edgeworthovom rozvoji pravdepodobnosti pokrytia štandardného intervalu pre odlišné nominálne koeficienty spoľahlivosti. Pre $\alpha = 0,05$ (pevná čiara), $\alpha = 0,01$ (prerušovaná) a $\alpha = 0,2$ (bodkovaná)

fakt, že pre pevné $n = 30$ 95%-ný interval dosahuje pokrytie cez 0,05, ale neexistuje také p , ktoré pokrýva interval s nominálnym 99%-ným koeficientom spoľahlivosti s pravdepodobnosťou aspoň 0,99.

4.3 Asymptotická dĺžka intervalov

Edgeworthov rozvoj druhého rádu nám v predchádzajúcej časti umožnil ucelené a jednotné porovnávanie intervalov spoľahlivosti v ich pravdepodobnosti pokrytia. Avšak v probléme intervalového odhadu neznámeho parametra je pokrytie iba jedna časť hodnotenia intervalov spoľahlivosti. Úspornosť (parsimony) prirodzene vyjadrená dĺžkou intervalu, je ďalšie dôležité kritérium posudzovania. Preto v tejto časti pre intervaly spomínané v kapitole 3 zavádzame asymptotický rozvoj ich strednej dĺžky. Uvažujeme asymptotický rozvoj do rádu $n^{-3/2}$. Tendenciou každého vhodného intervalu je byť čo najúspornejší, teda mať najmenšiu dĺžku. Toto kritérium nesmieme však brať za rozhodujúce. Váha tohoto kritéria býva obmedzená v situácii, keď sa intervaly značne líšia vo veľkosti pravdepodobnosti pokrytia. Teoretické výpočty sú tak trochu viac technické ako by sme si želali, podrobnejšie odvodenie uvádzam v dodatku. Avšak závery, ktoré môžeme získať, sú zrozumiteľné a prehľadné. Asymptotický rozvoj poukazuje na jednotnú štruktúru alternatívnych intervalov.

V tejto časti sa znovu zaoberáme strednou dĺžkou intervalov spoľahlivosti, avšak v tomto prípade sa zameriavame skôr na teoretické vlastnosti. Uvedme

preto definíciu strednej dĺžky pre $CI(X) = [L(X, n), U(X, n)]$ ešte raz

$$\mathbb{E}\{CI(X)\} = \sum_{x=0}^n (U(x, n) - L(x, n)) \binom{n}{x} p^x (1-p)^{n-x}.$$

Rozvoj pre strednú dĺžku sa kvalitatívne líši od Edgeworthovho rozvoja pre pokrytia v tom, že Edgeworthov rozvoj pre pravdepodobnosť pokrytia zahŕňa členy $n^{-1/2}$ a n^{-1} , zatiaľ čo rozvoj pre dĺžku zahŕňa členy rádu $n^{-1/2}$ a $n^{-3/2}$. Koeficienty u člena $n^{-1/2}$ sú rovnaké pre všetky intervaly a líšia sa u člena $n^{-3/2}$. Preto prirodzene za základ pre porovnanie dĺžky sa považuje koeficient u $n^{-3/2}$.

Veta 4.7 *Nech CI je všeobecné označenie pre každý z intervalov CI_S , CI_W , CI_{AC} , CI_J , CI_{LR} . Potom*

$$\mathbb{E}_{n,p}\{\text{dĺžka } CI\} = 2\kappa\sqrt{pq}n^{-1/2} \left(1 - \frac{\delta(\kappa, p)}{8npq}\right) + O(n^{-2}), \quad (4.30)$$

kde

$$\delta(\kappa, p) = 1 \quad \text{pre } CI_S, \quad (4.31)$$

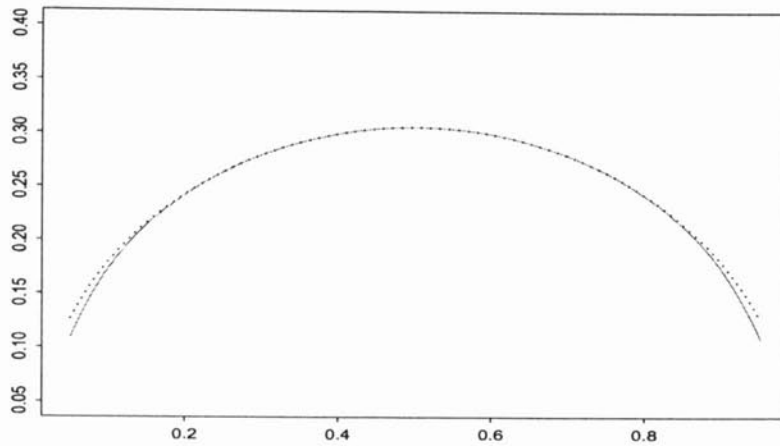
$$\delta(\kappa, p) = 1 + \kappa^2(8pq - 1) \quad \text{pre } CI_W, \quad (4.32)$$

$$\delta(\kappa, p) = 1 + \kappa^2(12pq - 2) \quad \text{pre } CI_{AC}, \quad (4.33)$$

$$\delta(\kappa, p) = 1 + \kappa^2 \left(\frac{26}{9}pq - \frac{2}{9}\right) \quad \text{pre } CI_{LR}, \quad (4.34)$$

$$\delta(\kappa, p) = 1 + \kappa^2 \left(\frac{26}{9}pq - \frac{2}{9}\right) + \frac{2}{9}(17pq - 2) \quad \text{pre } CI_J. \quad (4.35)$$

Dôkaz vety 4.7 je uvedený v dodatku. Uvedený asymptotický rozvoj (4.30) je dostatočne presný. Napríklad obrázok 4.8 zobrazuje odchýlku v aproximácii strednej dĺžky pre štandardný interval. Vidíme, že pre $0,1 \leq p \leq 0,9$ je táto aproximácia dostatočne presná. Porovnávanie koeficientov u člena $n^{-3/2}$ vedie k porovnávaniu výrazov $\delta(\kappa, p)$ pre jednotlivé intervaly. Najprv poukážeme na to, že pre všetky $\kappa > 0$ a všetky $0 \leq p \leq 1$ je $1 + \kappa^2(8pq - 1) \geq 1 + \kappa^2(12pq - 2)$, čo dokazuje fakt, že CI_{AC} obsahuje vždy CI_W ako podinterval, a preto je vždy dlhší. Pre iné páry intervalov porovnanie ich strednej dĺžky závisí na κ a p . Pre $\alpha = 0,05$, $\kappa = 1,96$, dostávame nasledujúce výsledky. Štandardný interval CI_S je najkratší pre $0 \leq p \leq 0,084$ a pre $0,916 \leq p \leq 1$. Interval založený na pomere vierohodnosti CI_{LR} je najkratší pre $0,084 \leq p \leq 0,137$ a pre $0,863 \leq p \leq 0,916$. Jeffreysov interval



Obrázok 4.8: Rozdiel medzi strednou dĺžkou (pevná čiara) a jej asymptotickým rozvojom (bodkovaná čiara) pre štandardný interval s $n = 40$ a $\alpha = 0,05$

CI_J je najkratší pre $0,137 \leq p \leq 0,201$ a pre $0,799 \leq p \leq 0,864$. A nakoniec Wilsonov interval CI_W je najkratší pre $0,201 \leq p \leq 0,799$. Na základe týchto výsledkov môžeme usúdiť, že Wilsonov interval je najkratší pre najväčšie rozpätie hodnôt p . Tieto závery opticky potvrdzuje aj obrázok 4.9, ktorý zobrazuje hodnoty $\delta(\kappa, p)$, tak ako sú definované vo vete 4.7. Obrázok porovnáva intervaly pre dve rôzne hodnoty n . Z toho môžeme pozorovať, že rozdiely medzi alternatívnymi intervalmi v ich strednej dĺžke (s presnosťou rádu $n^{-3/2}$) sú kvalitatívne rovnaké pre rôzne hodnoty n , aj keď čo to do veľkosti sú rozdielne. Z vyjadrenia asymptotického rozvoja strednej dĺžky vo vete 4.7 vidíme, že usporiadanie intervalov je rovnomerné pre všetky n a všetky κ . Tiež vidíme, že štandardný interval pre bežné hodnoty p nie je ani zďaleka najkratší. Tento fakt zdôrazňuje skutočnosť, že nedostatok štandardnej metódy nespočíva v tom, že interval je príliš krátky. Interval CI_S nepovažujeme za spoľahlivú voľbu, a to pre jeho veľmi slabé pokrytie. Medzi alternatívnymi metódami Jeffreysov interval a interval založený na verohodnostnom pomere sú najúspornejšie pre hraničné hodnoty p , inak je najúspornejší Wilsonov interval.

Analýza strednej dĺžky v binomickom rozdelení nám nedáva rovnaké hodnotenie pre všetky p , a tak kvôli všeobecnejším záverom platným pre všetky p uvažujeme tiež integrovanú verziu asymptotického rozvoja strednej dĺžky intervalov. Integrujeme cez celý parametrický priestor $(0, 1)$. Alternatívne by sme mohli uvažovať supremum strednej dĺžky.

Lemma 4.8 *Pre strednú dĺžku intervalov spoľahlivosti $CI_S, CI_W, CI_{AC}, CI_{LR}, CI_J$ definovaných vo vete 4.7 platí*

$$(i) \quad \int_0^1 E_{n,p}\{dĺžka(CI_S)\} dp = \frac{\kappa\pi}{4}n^{-1/2} - \frac{\kappa\pi}{4}n^{-3/2} + O(n^{-2}), \quad (4.36)$$

$$(ii) \quad \int_0^1 E_{n,p}\{dĺžka(CI_W)\} dp = \frac{\kappa\pi}{4}n^{-1/2} - \frac{\kappa\pi}{4}n^{-3/2} + O(n^{-2}), \quad (4.37)$$

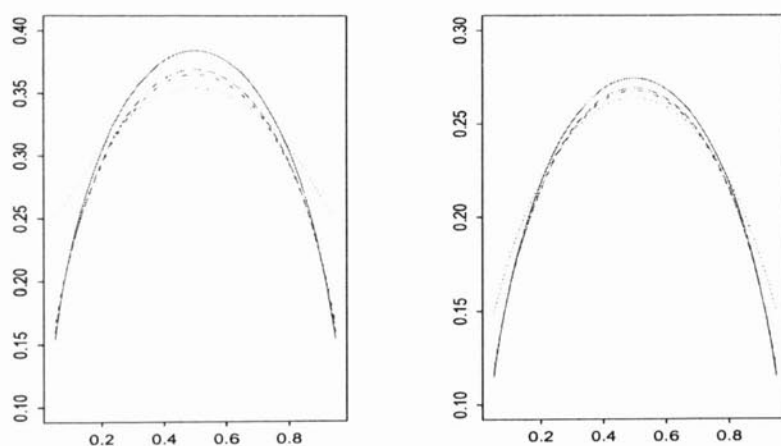
$$(iii) \quad \int_0^1 E_{n,p}\{dĺžka(CI_{AC})\} dp = \frac{\kappa\pi}{4}n^{-1/2} + \left(\frac{\kappa^2}{2} - 1\right) \frac{\kappa\pi}{4}n^{-3/2} + O(n^{-2}), \quad (4.38)$$

$$(iv) \quad \int_0^1 E_{n,p}\{dĺžka(CI_{LR})\} dp = \frac{\kappa\pi}{4}n^{-1/2} - \left(1 + \frac{5\kappa^2}{36}\right) \frac{\kappa\pi}{4}n^{-3/2} + O(n^{-2}), \quad (4.39)$$

$$(v) \quad \int_0^1 E_{n,p}\{dĺžka(CI_J)\} dp = \frac{\kappa\pi}{4}n^{-1/2} - \left(\frac{37}{36} + \frac{5\kappa^2}{36}\right) \frac{\kappa\pi}{4}n^{-3/2} + O(n^{-2}). \quad (4.40)$$

Dôkaz: Dôkaz plynie priamo z výrazov (4.31)–(4.35) integrovaním cez p . Použijeme vzťah $\int_0^1 \sqrt{p(1-p)} dp = \frac{1}{8}\pi$. ■

Už v prvej časti tejto kapitoly sme vyhodnocovali intervaly spoľahlivosti podľa kritéria koeficient strednej dĺžky. Ukázali sme, že integrovaná dĺžka alternatívnych intervalov rastie v poradí $CI_J, CI_{LR}, CI_W, CI_{AC}$. Tie sme porovnávali strednú dĺžku Wilsonovho a štandardného intervalu. Lemma (4.8) ukazuje, že rádovo s chybou n^{-2} rozdiel medzi intervalmi zaniká a integrovaná asymptotická dĺžka týchto dvoch intervalov je vždy identická. Porovnaním koeficientov u člena $n^{-3/2}$ získame zoradenie intervalov od najkratšieho po najdlhší v poradí $CI_J, CI_{LR}, CI_W, CI_S$ a CI_{AC} . Vidíme, že CI_J je vždy najmenší a CI_{AC} je vždy najväčší, nezávisle na hodnote κ , čo je hlavným prínosom tejto lemy. Stredná dĺžka je druhotné kritérium v posudzovaní intervalov spoľahlivosti. Avšak teoretickou analýzou prevedenou s chybou rádovo n^{-2} môžeme získať pekné závery. Znovu sa potvrdzuje nespoľahlivosť štandardnej metódy, pretože tento interval je buď príliš dlhý alebo príliš krátky, čo nie je pri intervalovom odhade parametra žiadateľné. Podobný záver platí aj pre Agresti-Coullov interval. Jeffreysov prístup a prístup založený na virohodnostnom pomere dávajú veľmi solídne výsledky a v zmysle integrovanej strednej dĺžky sú v danom poradí najkratšie.



Obrázok 4.9: Porovnanie asymptotického rozvoja strednej dĺžky pre Waldov (pevná čiara) interval, Wilsonov (prerušovaná), Agresti-Coullov (bodkovaná), Jeffreysov (bodkočiarkovaná) a interval založený na LR (dlhé pomlčky) s $n = 25$ (naľavo) a $n = 50$ (napravo)

Záverečné poznámky

Konštrukcia intervalov spoľahlivosti pre parameter p binomického rozdelenia patrí medzi veľmi časté aplikácie v matematickej štatistike. Často sa doporučuje tzv. štandardný Waldov interval, ktorý má takmer univerzálne uplatnenie. V tejto práci sa zaoberáme niektorými skutočnosťami súvisiacimi s touto témou. Hneď potom, čo stručne zhrnieme základné princípy konštruovania intervalov spoľahlivosti pre všeobecný parameter a spomenieme dualitu medzi intervalmi spoľahlivosti a testovaním hypotéz, venujeme sa štandardnému intervalu. Poukazujeme na jeho slabé a neuspokojivé vlastnosti. Na mnohých príkladoch sme ukázali, že obmedzenia, ktoré sa spájajú s používaním štandardného intervalu v praxi, nie sú platné. Štandardný interval trpí tak vážnymi nedostatkami, že by nemal byť vôbec používaný. Na základe týchto skutočností sa vynára potreba nových intervalov spoľahlivosti, čo nás v ďalšom viedlo k hľadaniu lepších alternatív. Tieto intervaly podrobujeme kompletnej numerickej i teoretickej analýze. Hoci priamo sa zameriavame len na štúdium štyroch odporúčaných intervalov, letmo predstavíme aj iné princípy konštruovania intervalového odhadu. Intervaly hodnotíme na základe navrhnutých nástrojov. Hoci empirické štúdium intervalov v ich pravdepodobnosti pokrytia a dĺžke je pre čitateľa veľmi efektné, musíme poznamenať, že výber hodnôt n , p a α v grafoch a tabuľkách je subjektívne a môže byť predmetom dohád. A tak pokračujúc v skúmaní intervalov po numerickej analýze intervaly študujeme i z teoretického hľadiska. Nástroje, ktoré používame, a to Edgeworthov rozvoj pravdepodobnosti pokrytia a asymptotický rozvoj dĺžky sú pracné a algebraicky obsiahle, avšak výsledky, ktoré nám poskytujú, sú zreteľné, jasné a uspokojivé. Osobitne veľmi významná je skutočnosť, že všetky teoretické závery, ktoré sú tu prezentované, sú v súlade s numerickými zisteniami a tým podporujú správnosť numerickej analýzy.

Aby sme to zhrnuli, záver je, že Agresti-Coullov interval má najväčšiu pravdepodobnosť pokrytia medzi ostatnými alternatívami, ale je tiež najdlhší a v priemere je o čosi viac konzervatívnejší, najmä ak p je blízko 0 alebo 1. Naproti tomu štandardný interval trpí systematickým negatívnym vychýlením v pravdepodobnosti pokrytia a jeho pokrytia je tak neuspokojivé. Slabé

vlastnosti štandardného intervalu však nutne nevyplývajú z toho, že by bol aj najúspornejší v zmysle dĺžky intervalu. Wilsonov interval, Jeffreysov interval a interval založený na vierohodnostnom pomere sú porovnateľné, s tým, že posledné dva intervaly sú víťazmi v tom zmysle, že takmer nulujú systematické vychýlenie v pravdepodobnosti pokrytia. Jeffreysov interval je v priemere najkratší. V dôsledku týchto zistení môžeme skonštatovať, že výber najvhodnejšieho intervalu spomedzi tu prezentovaných záleží na osobnej vôli a chuti.

V súvislosti s poslednou poznámkou, by som chcela zdôrazniť, že hoci sa v tejto práci zaoberáme hlavne intervalmi spoľahlivosti, ktorých pravdepodobnosť pokrytia spadá pre niektoré hodnoty p pod nominálny koeficient spoľahlivosti, všeobecne rozoznávame tri situácie, v ktorých sa štatistickí môžu rozhodovať pre vhodnú metódu. Jedna situácia je, keď potrebujeme garantovať pravdepodobnosť pokrytia najmenej rovnú nominálnemu koeficientu spoľahlivosti. Typicky však tieto intervaly majú pravdepodobnosť pokrytia o dosť väčšiu a sú dlhšie. Pokladá sa takýto interval spoľahlivosti nevyhnutne aj za lepší a bezpečnejší? Príliš veľký interval môže priamo v praxi viesť k rozhodnutiu o nečinnosti, čo nie je to, čo očakávame. Druhá situácia je tá, ktorá sa spája s intervalmi tu prezentovanými. Chceme, aby pravdepodobnosť pokrytia bola blízko nominálnemu koeficientu spoľahlivosti, ale nemusí byť nevyhnutne nad. Táto situácia je významnejšia pre štatistickú prax. Avšak tieto intervaly zlyhávajú v zmysle klasickej teoretickej definície „95%-ného intervalu spoľahlivosti“. Na základe výsledkov by tu naše odporúčanie viedlo k používaniu Jeffreysovho intervalu alebo intervalu založeného na vierohodnostnom pomere. Všetky štyri odporúčané intervaly však môžeme uprednostniť pred použitím štandardnej metódy. A tretia situácia sa spája s metódami, ktoré sú vhodné zaradiť do vyučovacieho procesu, kde názornosť a jednoduchosť prezentovaného intervalu sa často podriaďuje kvalite odhadu. V tejto situácii tak môžeme odporučiť Agresti-Coullov interval, hoci je trochu dlhší a trochu konzervatívnejší. Ak by sme chceli použiť štatistické možnosti softwaru, v programme R sú priamo implementované tieto metódy: štandardná, Wilsonova a exaktná metóda.

Problém intervalového odhadu pre parameter p binomického rozdelenia patrí medzi základné problémy v štatistike a hoci sa zdá, že je jednoduchý, je značne rozsiahly. V tejto práci som so obmedzila len na konštruovanie obojstranných 95%-ných intervalov spoľahlivosti. Hoci tu uvedené výsledky ostávajú kvalitatívne v platnosti i pre iné hodnoty α , problém konštruovania jednostranných intervalov spoľahlivosti sa značne líši od obojstranného problému, napriek niektorým spoločným znakom. Ďalej chcem tiež poznamenať, že problém intervalového odhadu v zmysle neželateľnej oscilácie sa prenáša aj do iných rozdelení s diskretným charakterom. Výsledky a metódy

tu prezentované môžeme aplikovať na iné diskkrétne rozdelenia ako negatívne binomické alebo Poissonovo. Napríklad rovnaké porovnanie platí pre skórový, Waldov a exaktný intervalový odhad pre Poissonov parameter λ , kde Waldov interval je založený na štatistike $(X - \lambda_0)/\sqrt{X}$, skórový na štatistike $(X - \lambda_0)/\sqrt{\lambda_0}$ a hranice exaktného intervalu sú $\frac{1}{2}(\chi_{0,025;2X}^2 \chi_{0,975;2(X+1)}^2)$. A práve v tomto smere by som videla prípadné rozšírenie tejto práce.

Na záver by som ešte rada dodala, že písanie tejto práce bolo pre mňa novou a zaujímavou skúsenosťou, príležitosťou obohatiť sa, či už poznaním nových oblastí v štatistike alebo nabraním ďalších skúseností s prácou v programe R a L^AT_EX. V neposlednom rade podnietila vo mne záujem o čítanie zaujímavých odborných článkov. Na tomto mieste patrí moje poďakovanie predovšetkým pánovi prof. RNDr. Jiřímu Andělovi, DrSc. Bez jeho pomoci by táto práca nikdy nevznikla. Budem spokojná, ak si tento text nájde v budúcnosti čitateľa a prispeje k jeho uznaniu, že štandardný interval trpí vážnymi nedostatkami v mnohých ohľadoch a že by vo všeobecnosti nemal byť používaný.

Dodatok A

Dôkazy a odvodenia

A.1 Aproximácia kvantilovej funkcia beta rozdelenia

Veta A.1 *Nech*

$$I_x(a, b) = \frac{1}{B(a, b)} \int_0^x t^{a-1} (1-t)^{b-1} dt \quad 0 \leq x \leq 1$$

je neúplná beta funkcia a nech $\Phi(y)$ je distribučná funkcia štandardného normálneho rozdelenia. Zavedme x_p a y_p tak, že $I_{x_p}(a, b) = p$ a $\Phi(y_p) = p$. Nech

$$h = 2 \left(\frac{1}{2a-1} + \frac{1}{2b-1} \right)^{-1}, \quad \lambda = \frac{y_p^2 - 3}{6},$$

$$\omega = \frac{y_p(h + \lambda)^{1/2}}{h} - \left(\frac{1}{2b-1} - \frac{1}{2a-1} \right) \left(\lambda + \frac{5}{6} - \frac{2}{3h} \right).$$

Potom platí

$$x_p \approx \frac{a}{a + b \cdot e^{2\omega}}.$$

Dôkaz: Dôkaz môžeme nájsť v Abramowitz a Stegun (1970) [5], uvedený na strana 269. ■

A.2 Edgeworthov rozvoj

Edgeworthov rozvoj je veľmi účinný nástroj pri študovaní vlastností pokrytia intervalov spoľahlivosti, ale na druhej strane je veľmi nepraktický pre zdĺhavé

algebraické úpravy. V tejto časti sú uvedené hlavné dôkazy a odvodenie z kapitoly 4. Binomické rozdelenie patrí medzi rozdelenia s diskretnou štruktúrou. Predpis Edgeworthovho rozvoja distribučnej funkcie pre triedu diskretných rozdelení je uvedená v Bhattacharya a Rao [5].

Lemma A.2 *Nech $X \sim \text{Bi}(n, p)$ a $\hat{p} = X/n$. Označme $S_n = n^{1/2}(\hat{p} - p)/\sqrt{pq}$ a $F_n(z) = P(S_n \leq z)$. Definujme funkciu $g(p, z) = g(p, z, n)$ ako v (4.13). Potom*

$$\begin{aligned} F_n(z) = & \Phi(z) + \left(\frac{1}{6} - \frac{p}{3}\right) (1 - z^2) \phi(z) \frac{1}{\sqrt{pq}} n^{-1/2} \\ & + \left(\frac{1}{2} - g(p, z)\right) \phi(z) \frac{1}{\sqrt{pq}} n^{-1/2} \\ & + \left\{ \left(\frac{pq}{18} - \frac{1}{72}\right) \kappa^5 + \left(\frac{7}{72} - \frac{11pq}{36}\right) \kappa^3 + \left(\frac{pq}{12} - \frac{1}{12}\right) \kappa \right\} \phi(z) \frac{1}{pq} n^{-1} \\ & + \left\{ \left(\frac{1}{6} - \frac{p}{3}\right) (z^3 - 3) \left(\frac{1}{2} - g(p, z)\right) \right. \\ & \left. - \left(\frac{1}{2} g^2(p, z) - \frac{1}{2} g(p, z) + \frac{1}{6}\right) \right\} z \phi(z) \frac{1}{pq} n^{-1} + O(n^{-3/2}). \end{aligned} \quad (\text{A.1})$$

Nech $z = z(n)$ závisí na n a nech môžeme z vyjadriť ako

$$z = \lambda_1 + \lambda_2 n^{-1/2} + \lambda_3 n^{-1} + O(n^{-3/2})$$

kde λ_1, λ_2 a λ_3 sú konštanty. Potom

$$\begin{aligned} F_n(z) = & \Phi(\lambda_1) + \left(\lambda_2 \sqrt{pq} + \frac{1}{6}(1 - 2p)(1 - \lambda_1^2)\right) \phi(\lambda_1) \frac{1}{\sqrt{pq}} n^{-1/2} \\ & + \left(\frac{1}{2} - g(p, z)\right) \phi(\lambda_1) \frac{1}{pq} n^{-1/2} \\ & + \left\{ \lambda_3 - \frac{1}{2} \lambda_1 \lambda_2^2 + \frac{1}{6}(1 - 2p) \frac{1}{\sqrt{pq}} \lambda_1 \lambda_2 (\lambda_1^2 - 3) \right\} \phi(\lambda_1) n^{-1} \\ & + \left\{ (4pq - 1) \lambda_1^5 + (7 - 22pq) \lambda_1^3 + (6pq - 6) \lambda_1 \right\} \phi(\lambda_1) \frac{1}{\sqrt{72pq}} n^{-1} \\ & + \left\{ \left[\frac{1}{6}(1 - 2p)(\lambda_1^2 - 3) - (pq)^{1/2} \lambda_2 \right] \left(\frac{1}{2} - g(p, z)\right) \right. \\ & \left. - \left[\frac{1}{2} g^2(p, z) - \frac{1}{2} g(p, z) + \frac{1}{12} \right] \right\} \lambda_1 \phi(\lambda_1) \frac{1}{pq} n^{-1} + O(n^{-3/2}). \end{aligned} \quad (\text{A.2})$$

Dôkaz: Rozvoj (A.1) plynie priamo z Vety 23.1 v Bhattacharya a Rao (1976) [5]. Ak $z = \lambda_1 + \lambda_2 n^{-1/2} + \lambda_3 n^{-1} + O(n^{-3/2})$, rozvinieme $\Phi(z)$ a $\phi(z)$ okolo λ_1 nasledovne

$$\Phi(z) = \Phi(\lambda_1) + \lambda_2 \phi(\lambda_1) n^{-1/2} + \left(\lambda_3 - \frac{1}{2} \lambda_1 \lambda_2^2 \right) \phi(\lambda_1) n^{-1} + O(n^{-3/2}), \quad (\text{A.3})$$

$$\phi(z) = \phi(\lambda_1) - \lambda_1 \lambda_2 \phi(\lambda_1) n^{-1/2} + O(n^{-1}). \quad (\text{A.4})$$

Teraz rozvoj (A.2) dostaneme dosadením (A.3) a (A.4) do (A.1). ■

Štandardný interval

Nech $X \sim \text{Bi}(n, p)$, $\hat{p} = X/n$ a $\hat{q} = 1 - \hat{p}$. Označme $\kappa = \Phi^{-1}(1 - \alpha/2)$. Potom $100(1 - \alpha)\%$ štandardný interval spoľahlivosti pre parameter p definujeme predpisom $CI_S = \hat{p} \pm \kappa^2 n^{-1/2} \sqrt{\hat{p}(1 - \hat{p})}$.

Dôkaz Vety 4.2: Zavedme značenie

$$\begin{aligned} A &= n + \kappa^2, \\ B &= n(2np + \kappa^2), \\ C &= n^3 p^2. \end{aligned}$$

Hľadáme l_S, u_S ktoré vyhovujú predpisu

$$P_S = \mathbb{P}_p(p \in CI_S) = \mathbb{P}(l_S \leq n^{1/2}(\hat{p} - p)/\sqrt{pq} \leq u_S).$$

Riešením kvadratickej nerovnice dostávame

$$(l_S, u_S) = \left(\frac{B \pm \sqrt{B^2 - AC}}{2A} - np \right) \frac{1}{\sqrt{pq}} n^{-1/2}, \quad (\text{A.5})$$

znamienko $+$ odpovedá u_S a znamienko $-$ odpovedá l_S . Asymptotický rozvoj výrazov l_S a u_S vedie k

$$(l_S, u_S) = \frac{\kappa^2(1/2 - p)}{\sqrt{npq}} \pm \left\{ \kappa + \frac{(1/8 - pq)\kappa^3}{npq} \right\} + O(n^{-3/2}). \quad (\text{A.6})$$

Označme

$$\begin{aligned} \lambda_2^S &= \kappa^2(1/2 - p)/\sqrt{pq}, \\ \lambda_3^S &= \kappa^3(1/8 - pq)/pq. \end{aligned}$$

Potom dosadením do (A.6) dostávame

$$\begin{aligned} l_S &= -\kappa + \lambda_2^S n^{-1/2} - \lambda_3^S n^{-1} + O(n^{-3/2}), \\ u_S &= \kappa + \lambda_2^S n^{-1/2} + \lambda_3^S n^{-1} + O(n^{-3/2}). \end{aligned}$$

Teraz môžeme pravdepodobnosť pokrytia $P_S = P_p(p \in CI_S) = F_n(u_S) - F_n(l_S)$ rozvinúť do Edgeworthovej rady využitím (A.2), čím sme odvodili (4.18). Úpravou z (A.6) dostávame nasledovný tvar hraníc štandardného intervalu spoľahlivosti

$$(l_S, u_S) = \frac{\kappa^2 n^{1/2} (1/2 - p) \pm \kappa n (pq + \kappa^2 / (4n))^{1/2}}{\sqrt{pq} (n + \kappa^2)}. \quad (\text{A.7})$$

■

Wilsonov interval

Dôkaz Vety 4.3: Edgeworthov rozvoj pre Wilsonov interval je jednoduchší, pretože

$$P_W = P_p(p \in CI_W) = P(-\kappa \leq n^{1/2}(\hat{p} - p)/\sqrt{pq} \leq \kappa).$$

Edgeworthov rozvoj vyplýva priamo z (A.2) dosadením $\lambda_1 = \pm\kappa$ za z . ■

Agresti-Coull interval

Dôkaz Vety 4.4: Dôkaz je podobný dôkazu pre štandardný interval. Označme

$$\begin{aligned} A &= n + 2\kappa^2, \\ B &= 2pn^2 + 4\kappa^2 pn + (2p - 1)\kappa^4, \\ C &= p^2 n^3 + \kappa^2 p(3p - 1)n^2 + \kappa^4 \left(3p^2 - 2p - \frac{1}{4}\right) n - pq\kappa^6. \end{aligned}$$

Riešením kvadratickej nerovnice

$$P_{AC} = P_p(p \in CI_{AC}) = P(l_{AC} \leq n^{1/2}(\hat{p} - p)/\sqrt{pq} \leq u_{AC})$$

nájdem l_{AC} , u_{AC} vyhovujúce

$$(l_{AC}, u_{AC}) = \left(\frac{B \pm \sqrt{B^2 - AC}}{2A} - np \right) \frac{1}{\sqrt{pq}} n^{-1/2}.$$

Znamienko $+$ patrí u_{AC} a znamienko $-$ patrí l_{AC} . Asymptotický rozvoj l_{AC} a u_{AC} vedie k

$$(l_{AC}, u_{AC}) = \pm \left\{ \kappa + \left(\frac{1}{8pq} - \frac{1}{2} \right) \kappa^3 n^{-1} \right\} + O(n^{-3/2}). \quad (\text{A.8})$$

Edgeworthov rozvoj pre Agresti-Coull interval (4.20) teraz vyplýva z rovnosti $P_{AC} = F_n(u_{AC}) - F_n(l_{AC})$ a z (A.2). ■

Interval založený na pomere vierohodnosti

Nech $X \sim \text{Bi}(n, p)$, označme $\hat{p} = X/n$ a $\hat{q} = 1 - \hat{p}$. Vierohodnostný pomer pre parameter p definujeme ako

$$\Lambda_n = \frac{p_0^X (1 - p_0)^{n-X}}{(X/n)^X (1 - X/n)^{n-X}},$$

kde X/n je maximálne vierohodný odhad parametra p .

Dôkaz Vety 4.6: V súlade so zavedením označením vierohodnostný pomer Λ_n môžeme prepísať do tvaru

$$\Lambda_n = \left(\frac{p}{\hat{p}}\right)^{n\hat{p}} \left(\frac{q}{\hat{q}}\right)^{n\hat{q}}$$

Nech $z = \sqrt{n}(\hat{p} - p)/\sqrt{pq}$, potom algebraickými úpravami dostávame, že nerovnosť $-2 \log \Lambda_n \leq \kappa^2$ je ekvivalentná s

$$p(1 + (np)^{-1/2}q^{1/2}z) \log(1 + (np)^{-1/2}q^{1/2}z) + \\ q(1 + (nq)^{-1/2}p^{1/2}z) \log(1 + (nq)^{-1/2}p^{1/2}z) - \frac{\kappa^2}{2n} \leq 0. \quad (\text{A.9})$$

Považujme pravú stranu nerovnosti (A.9) za funkcie v premennej z , označme ju $v(z)$. Ľahko sa dá overiť, že je konvexná funkcia, a tak má najviac dva korene. Označme l_{LR} a u_{LR} korene rovnice $v(z) = 0$. Teda platí

$$v(l_{LR}) = v(u_{LR}) = 0.$$

V ďalšom potrebujeme explicitné vyjadrenie l_{LR} a u_{LR} . Taylorovým rozvojom rozvineme funkciu $b(t) = (1 + t) \log(1 + t)$ do rady

$$b(t) = t + \frac{1}{2}t^2 - \frac{1}{6}t^3 + \frac{1}{12}t^4 + O(t^5). \quad (\text{A.10})$$

Teraz aplikujeme (A.10) na (A.9) a algebraickými úpravami dostaneme ekvivalentný prepis rovnosti $v(z) = 0$ na

$$z^2 - \frac{1}{3\sqrt{pq}}(1 - 2p)n^{-1/2}z^3 + \frac{1}{6pq}(1 - 3pq)n^{-1}z^4 - \kappa^2 = O(n^{-3/2}). \quad (\text{A.11})$$

Položme $z = \pm\kappa + b_1n^{-1/2} + b_2n^{-1}$. Dosadení do rovnosti (A.11) a úpravou dostávame

$$\left(2b_1\kappa - \frac{1}{3\sqrt{pq}}(1 - 2p)n^{-1/2}\kappa^3\right)n^{-1/2} +$$

$$\left(\pm 2b_2\kappa + b_1^2 - \frac{1}{3\sqrt{pq}}(1-2p)3b_1\kappa^2 + \frac{1}{6pq}(1-3pq)\kappa^4 \right) n^{-1} = O(n^{-3/2}).$$

Nájdeme riešenie pre b_1 a b_2

$$b_1 = \frac{1}{6\sqrt{pq}}(1-2p)\kappa^2 \quad \text{a} \quad b_2 = \mp \frac{1}{72} \left(\frac{1}{pq} + 2 \right) \kappa^3.$$

Po zanedbaní člena $O(n^{-3/2})$ sme našli korene rovnice $v(z) = 0$

$$(l_{LR}, u_{LR}) = \frac{1}{6\sqrt{pq}}(1-2p)\kappa^2 n^{-1/2} \pm \left\{ \kappa - \frac{1}{72} \left(\frac{1}{pq} + 2 \right) \kappa^3 n^{-1} \right\} \quad (\text{A.12})$$

Pravdepodobnosť pokrytia intervalu založeného na vierohodnostnom pomere teraz vyhovuje vzťahu

$$P_p(p \in CI_{LR}) = P(l_{LR} \leq n^{1/2}(\hat{p} - p)/\sqrt{pq} \leq u_{LR}).$$

Ekvivalentne môžeme písať $P_{LR} = F(u_{LR}) - F_n(l_{LR})$ a Edgeworthov rozvoj pre interval založený na vierohodnostnom pomere (4.22) vyplýva z (A.2). ■

Jeffreysov interval

Nech $X \sim \text{Bi}(n, p)$. Potom $100(1 - \alpha)\%$ Jeffreysov interval spoľahlivosti pre parameter p definujeme predpisom

$$\begin{aligned} CI_J &= [p_l, p_u] \\ &= [\beta(\alpha/2; X + 1/2, n - X + 1/2), \beta(1 - \alpha/2; X + 1/2, n - X + 1/2)]. \end{aligned}$$

kde $\beta(\alpha; m_1, m_2)$ označujú kvantily $B(m_1, m_2)$ rozdelenia.

Dôkaz Vety 4.5 Označme $F(z; m_1, m_2)$ distribučnú funkciu $B(m_1, m_2)$ rozdelenia. Potom pre pravdepodobnosť pokrytia parametra p platí

$$\begin{aligned} P_p(p \in CI_J) &= \\ &= P\{\beta(\alpha/2; X + 1/2, n - X + 1/2) \leq p \leq \beta(1 - \alpha/2; X + 1/2, n - X + 1/2)\} \\ &= P\{\alpha/2 \leq F(p; X + 1/2, n - X + 1/2) \leq 1 - \alpha/2\} \end{aligned}$$

Pre dané p , n je distribučná funkcia $F(p; X + 1/2, n - X + 1/2)$ je rastúca v X . A tak existuje práve jedno riešenie $X_l = \rho_1(1 - \alpha/2, p)$, resp. $X_u = \rho_2(\alpha/2, p)$ splňujúce

$$\begin{aligned} F(p; X_l + 1/2, n - X_l + 1/2) &\leq 1 - \alpha/2 \quad \text{a} \\ F(p; X_l - 1 + 1/2, n - (X_l - 1) + 1/2) &> 1 - \alpha/2, \end{aligned}$$

$$F(p; X_u + 1/2, n - X_u + 1/2) \geq \alpha/2 \quad \text{a} \\ F(p; X_u + 1 + 1/2, n - (X_u + 1) + 1/2) < \alpha/2.$$

Pravdepodobnosť pokrytia pre Jeffreysov interval môžeme teraz vyjadriť ako

$$P_J = \mathbb{P}_p(p \in CI_J) = \mathbb{P}(l_J \leq n^{1/2}(\hat{p} - p)/\sqrt{pq} \leq u_J),$$

s l_J a u_J danými nasledovným predpisom

$$l_J = [\rho_1(1 - \alpha/2, p) - np] \frac{1}{\sqrt{pq}}, \\ u_J = [\rho_2(\alpha/2, p) - np] \frac{1}{\sqrt{pq}}. \quad (\text{A.13})$$

Kvantily v (A.13) sú definované implicitne cez ρ_1, ρ_2 . Pre dôkaz vety (4.5) potrebujeme nájsť asymptotický rozvoj pre l_J, u_J

$$(l_J, u_J) = \pm \kappa + \left[\left(\frac{1}{6} \kappa^2 - \frac{1}{6} \right) + \left(\frac{1}{3} - \frac{1}{3} \kappa^2 \right) p \right] \frac{1}{\sqrt{pq}} n^{-1/2} \\ \pm \left\{ (8pq)^{-1} \kappa^3 + \left(\frac{1}{3} - \frac{1}{3} \kappa^2 \right) \kappa + r_2(p) \frac{1}{\sqrt{pq}} \right. \\ \left. - \left(\frac{1}{2} - p \right) \frac{1}{pq} r_1(p) \kappa \right\} n^{-1} + O(n^{-3/2}), \quad (\text{A.14})$$

kde

$$r_1(p) = \left(\frac{1}{3} \kappa^2 + \frac{1}{6} \right) (1 - 2p), \quad (\text{A.15})$$

$$r_2(p) = \left[\left(\frac{1}{36} \kappa^2 + \frac{1}{18} \right) \frac{1}{pq} - \left(\frac{13}{36} \kappa^2 + \frac{17}{36} \right) \right] \kappa \sqrt{pq}. \quad (\text{A.16})$$

Asymptotický rozvoj kvantilov l_J a u_J získame priamo z asymptotického rozvoja kvantilov $B(a, b)$ rozdelenia v [7]. Rozvoj pre Jeffreysov interval (4.21) teraz plyní z (A.2). ■

A.3 Asymptotický rozvoj dĺžky

Odvedenie rovností (4.31) -(4.35) je algebraicky veľmi pracné. Preto sú tu uvedené len hlavné kroky a medzikroky vedúce k zjednodušeniu výrazov niekedy preskočíme.

V celom dôkaze sa používa značenie $\omega = (X/n - p)/p$ a $\tau = p/q$. Rovnako zavedme značenie L (length) pre dĺžku všeobecného intervalu spoľahlivosti.

Dôkaz Vety 4.7:

Štandardný interval

Dĺžku štandardného intervalu definovaného v (2.3) v súlade so zavedením značením môžeme vyjadriť ako

$$L_S = 2\kappa n^{-1/2} \left[\frac{X}{n} \left(1 - \frac{X}{n} \right) \right]^{1/2} = 2\kappa n^{-1/2} \sqrt{pq} (1 + \omega)^{1/2} (1 + \tau\omega)^{1/2}.$$

Rozvojom $(1 + \omega)^{1/2} (1 - \tau\omega)^{1/2}$ do asymptotickej rady dostávame

$$L_S = 2\kappa n^{-1/2} \sqrt{pq} \left\{ 1 + \frac{1 - \tau}{2} \omega - \frac{(1 + \tau)^2}{8} \omega^2 + R(\omega) \right\}, \quad (\text{A.17})$$

kde $R(\omega)$ je zvyšok rady. Keďže platí $E|X - np|^3 = O(n^{3/2})$, zvyšok rady vyhovuje $E(|R(\omega)|) = O(n^{-3/2})$. Aplikovaním strednej hodnoty na (A.17) a s využitím $E(\omega) = 0$ dostávame

$$E(L_S) = 2\kappa \sqrt{pq} n^{-1/2} \left(1 - \frac{1}{8npq} \right) + O(n^{-2}),$$

čím sme tvrdenie (4.31) dokázali.

Wilsonov interval

Dĺžka Wilsonovho intervalu definovaného v (3.1) je daná ako

$$\begin{aligned} L_W &= 2\kappa n^{-1/2} \frac{n}{n + \kappa^2} \left[\frac{X}{n} \left(1 - \frac{X}{n} \right) + \frac{\kappa^2}{4n} \right]^{1/2} \\ &= 2\kappa n^{-1/2} \sqrt{pq} [1 - \kappa^2 n^{-1} + O(n^{-2})] \\ &\quad \left\{ 1 + \frac{1 - \tau}{2} \omega - \frac{(1 + \tau)^2}{8} \omega^2 + \frac{\kappa^2}{8npq} + R(\omega) \right\}, \end{aligned} \quad (\text{A.18})$$

kde zvyšok rady vyhovuje $E(|R(\omega)|) = O(n^{-3/2})$. Priamo z (A.18) plynie

$$\begin{aligned} E(L_W) &= 2\kappa n^{-1/2} \sqrt{pq} [1 - \kappa^2 n^{-1} + O(n^{-2})] \left(1 - \frac{\kappa^2 - 1}{8npq} + O(n^{-3/2}) \right) \\ &= 2\kappa n^{-1/2} \sqrt{pq} \left(1 + \frac{8\kappa^2 pq + 1 - \kappa^2}{8npq} \right) + O(n^{-2}), \end{aligned}$$

čím je (4.32) dokázané.

Agresti-Coullov interval

Odvozenie asymptotického rozvoja pre strednú dĺžku Agresti-Coullovho intervalu je podobné ako pre Wilsonov interval. Nech L_{AC} je dĺžka intervalu CI_{AC} definovaného v (3.5). Potom platí

$$\begin{aligned}
 L_{AC} &= 2\kappa(n + \kappa^2)^{-1/2} \left[\frac{X + \kappa^2/2}{n + \kappa^2} \left(1 - \frac{X + \kappa^2/2}{n + \kappa^2} \right) \right]^{1/2} \\
 &= 2\kappa n^{-1/2} \left(\frac{n}{n + \kappa^2} \right)^{-3/2} \left[\frac{X}{n} \left(1 - \frac{X}{n} \right) + \frac{\kappa^2}{2n} + \frac{\kappa^4}{4n^2} \right]^{1/2} \\
 &= 2\kappa n^{-1/2} \sqrt{pq} \left[1 - \frac{3}{2} \kappa^2 n^{-1} + O(n^{-2}) \right] \\
 &\quad \left\{ 1 + \frac{1 - \tau}{2} \omega - \frac{(1 + \tau)^2}{8} \omega^2 + \frac{\kappa^2}{4npq} + R(\omega) \right\}, \tag{A.19}
 \end{aligned}$$

kde pre zvyšok rady platí $E(|R(\omega)|) = O(n^{-3/2})$. Pre strednú hodnotu (A.19) platí

$$\begin{aligned}
 E(L_{AC}) &= 2\kappa n^{-1/2} \sqrt{pq} \left[1 - \frac{3}{2} \kappa^2 n^{-1} + O(n^{-2}) \right] \left(1 - \frac{1 - 2\kappa^2}{8npq} \right) + O(n^{-2}) \\
 &= 2\kappa n^{-1/2} \sqrt{pq} \left(1 + \frac{12\kappa^2 pq + 1 - 2\kappa^2}{8npq} \right) + O(n^{-2}),
 \end{aligned}$$

čím sme dokázali tvrdenie (4.33).

Interval založený na pomere vierohodnosti

Toto je najzložitejší prípad, pretože nepoznáme explicitné vyjadrenie koncových bodov intervalu. Nech $X \sim \text{Bi}(n, p)$, predpokladajme, že $\hat{p} = X/n > 0$, pomer vierohodnosti je daný

$$\begin{aligned}
 \Lambda_n &= \frac{p^X (1 - p)^{n-X}}{(X/n)^X (1 - X/n)^{n-X}} \\
 &= \left(\frac{p}{\hat{p}} \right)^{n\hat{p}} \left(\frac{q}{\hat{q}} \right)^{n\hat{q}}.
 \end{aligned}$$

Koncové body intervalu založeného na vierohodnostnom pomere sú korene rovnice $-\log \Lambda_n = \kappa^2/2$. Odvozenie asymptotického rozvoja spočíva v dvoch krokoch. V prvom kroku sa nájde asymptotický rozvoj týchto koreňov až do rádu $n^{-3/2}$ a potom sa nájde rozvoj pre strednú hodnotu ich rozdielu.

Označme $t = p/\hat{p} - 1$. Potom dostávame nasledovný tvar rovnice

$$\log(1+t) - \frac{\hat{q}}{\hat{p}} \log\left(1 - \frac{\hat{p}}{\hat{q}}t\right) = \kappa^2/(2n\hat{p}).$$

Asymptotické korene sa dajú nájsť rovnakým spôsobom ako pri odvodzovaní Edgeworthovho rozvoja v bodoch (A.9)–(A.12). Riešenie je uvedené v Brown, Cai a DasGupta (2002) [8]. Označme korene rovnice \underline{t} a \bar{t} . Potom máme

$$\underline{t} = -\kappa\sqrt{\hat{q}/\hat{p}}n^{-1/2} + \frac{1}{3}\kappa^2(1-2\hat{p})\frac{1}{\hat{p}}n^{-1} - \frac{1}{36}\kappa^3(1-13\hat{p}\hat{q})\frac{1}{\hat{p}\sqrt{\hat{q}\hat{p}}}n^{-3/2} + R_1 \quad (\text{A.20})$$

a

$$\bar{t} = \kappa\sqrt{\hat{q}/\hat{p}}n^{-1/2} + \frac{1}{3}\kappa^2(1-2\hat{p})\frac{1}{\hat{p}}n^{-1} + \frac{1}{36}\kappa^3(1-13\hat{p}\hat{q})\frac{1}{\hat{p}\sqrt{\hat{q}\hat{p}}}n^{-3/2} + R_2, \quad (\text{A.21})$$

kde $E(|R_1|) = O(n^{-2})$ a $E(|R_2|) = O(n^{-2})$. Z (A.20) a (A.21) pre dĺžka intervalu založeného na pomere vierohodnosti platí

$$L_{LR} = 2\kappa\sqrt{\hat{q}\hat{p}}n^{-1/2} + \frac{1}{18}\kappa^3(1-13\hat{p}\hat{q})\frac{1}{\sqrt{\hat{q}\hat{p}}}n^{-3/2} + O(n^{-2}). \quad (\text{A.22})$$

Dá sa ukázať ako v (A.17), že platí

$$E[(\hat{q}\hat{p})^{1/2}] = E\left[\left(\frac{X}{n}\left(1 - \frac{X}{n}\right)\right)^{1/2}\right] = \sqrt{pq}\left(1 - \frac{1}{8npq}\right) + O(n^{-3/2})$$

a tak z (A.22) plynie

$$\begin{aligned} E(L_{LR}) &= 2\kappa\sqrt{pq}\left(1 - \frac{1}{8npq}\right)n^{-1/2} + \frac{1}{18}\kappa^3(1-13pq)\frac{1}{\sqrt{pq}}n^{-3/2} + O(n^{-2}) \\ &= 2\kappa(pq)^{1/2}n^{-1/2}\left(1 - \frac{9 + \kappa^2(26pq - 2)}{72npq}\right) + O(n^{-2}), \end{aligned}$$

čím je (4.34) dokázané.

Jeffreysov interval

Aproximácia koncových body $100(1-\alpha)\%$ -ného Jeffreysovho intervalu je uvedená v Brown, Cai a DasGupta (2002) [8]. Nech $\hat{p} = X/n$ a $\hat{q} = 1 - \hat{p}$. Priamo z (60) v [8] máme

$$\begin{aligned} (L_J, U_J) &= \hat{p} + \left(\frac{1}{3}\kappa^2 + \frac{1}{6}\right)(1-2\hat{p})n^{-1} \pm \left\{ \kappa\sqrt{\hat{p}\hat{q}}n^{-1/2} \right. \\ &\quad \left. + \left[\left(\frac{1}{36}\kappa^2 + \frac{1}{18}\right)\frac{1}{\hat{p}\hat{q}} - \left(\frac{13}{36}\kappa^2 + \frac{17}{36}\right)\right] \kappa\sqrt{\hat{p}\hat{q}}n^{-3/2} \right\} + O(n^{-2}), \end{aligned}$$

znamienko $+$ patrí U_J a znamienko $-$ patrí L_J . Teraz môžeme dĺžku Jeffreysovho intervalu vyjadriť ako

$$\begin{aligned} L_J = 2\kappa n^{-1/2} \left[\frac{X}{n} \left(1 - \frac{X}{n} \right) \right]^{1/2} + 2 \left[\left(\frac{1}{36} \kappa^2 + \frac{1}{18} \right) \frac{1}{\hat{p}\hat{q}} \right. \\ \left. - \left(\frac{13}{36} \kappa^2 + \frac{17}{36} \right) \right] \kappa \sqrt{\hat{p}\hat{q}} n^{-3/2} + O(n^{-2}). \end{aligned} \quad (A.23)$$

Poznajmenajme, že prvý člen v (A.23) je priamo rovný dĺžke štandardného intervalu C_S . Preto priamo z (4.31) dostávame

$$\begin{aligned} E(L_J) &= 2\kappa \sqrt{pq} n^{-1/2} \left(1 - \frac{1}{8npq} \right) \\ &+ 2 \left[\left(\frac{1}{36} \kappa^2 + \frac{1}{18} \right) (\hat{p}\hat{q})^{-1} - \left(\frac{13}{36} \kappa^2 + \frac{17}{36} \right) \right] \kappa \sqrt{\hat{p}\hat{q}} n^{-3/2} + O(n^{-2}) \\ &= 2 \kappa \sqrt{pq} n^{-1/2} \left(1 - \frac{(2/9)(13\kappa^2 + 17)pq - (2/9)(\kappa^2 + 2)}{8npq} \right) + O(n^{-2}). \end{aligned}$$

Tým sme overili tvrdenie (4.35). Veta 4.7 je dokázaná. ■

Literatúra

- [1] Abramowitz M., Stegun I. A. (1970): *Handbook of Mathematical Functions*. Dover, New York.
- [2] Agresti A., Coull B. A. (1998): Approximate is better than “exact” for interval estimation of binomial proportion. *The American Statistician* **52**, 119–126.
- [3] Agresti A., Min Y. (2001): On small-sample confidence intervals for parameters in discrete distributions. *Biometrics* **57**, 963–971.
- [4] Anděl J. (2005): *Základy matematické statistiky*, 1. vydanie. Matfyzpress, Praha.
- [5] Bhattacharya R. N., Rao R. R. (1976): *Normal Approximation and Asymptotic expansions*. Wiley, New York.
- [6] Blyth C. R., Still H. A. (1983): Binomial confidence intervals. *Journal of American Statistical Association* **78**, 108–116.
- [7] Brown L. D., Cai T., DasGupta A. (2001): Interval estimation for a binomial proportion (with discussion). *Statistical Science* **16**, 101–133.
- [8] Brown L. D., Cai T., DasGupta A. (2002): Confidence interval for a binomial proportion and asymptotic expansions. *Annals of Statistics* **30**, 160–201.
- [9] Brown L. D., Cai T., DasGupta A. (2001): Interval estimation in exponential families. *Statistica Sinica* **13**, 19–49.
- [10] Clopper C. J., Pearson E. S. (1934): The use of confidence or fiducial limits illustrated in the case of binomial. *Biometrika* **26**, 404–413.
- [11] Edwardes M. D. (1998): The evaluation of confidence sets with application to binomial intervals. *Statistica Sinica* **8**, 393–409.

- [12] Ghosh B. K. (1979): On the relation among shortest confidence intervals of different types. *Calcutta Statist. Assoc. Bull.* **10**, 147–152.
- [13] Ghosh B. K. (1961): A comparison of some approximate confidence intervals for the binomial parameter. *Journal of the American Statistical Association* **74**, 894–900.
- [14] Hall P.(1992): *The Bootstrap and Edgeworth Expansion*, 2nd Edition. Springer-Verlag, New York.
- [15] Hušková M. (1985): *Bayesovské metody*, 1. vydanie. Univerzita Karlova, Praha.
- [16] Huwang L. (1995): A note on the accuracy of an approximate interval for the binomial parameter. *Statistics a Probability Letters* **24**, 177–180.
- [17] Chen H. (1990): The accuracy of approximate intervals for a binomial parameter. *Journal of the American Statistical Association* **85**, 514–518.
- [18] Jurečková J. (1982): *Testy parametrických hypotéz*, skripta. SPN, Praha.
- [19] Lehmann E. L. (1986): *Testing Statistical Hypotheses*, 2nd Edition. Wiley, New York.
- [20] Peizer D. B., Pratt J. W. (1968): A normal approximation for binomial, F, Beta, and other common, related tail probabilities, II. *Journal of the American Statistical Association* **63**, 1416–1456.
- [21] Pratt J. W. (1961): Length of confidence interval. *Journal of the American Statistical Association* **56**, 549–567.
- [22] Rao C. R. (1973): *Linear Statistical Inference and its Applications*, 2nd Edition. Wiley, New York.
- [23] Ulicny J. (2001): *A Comparison of Binomial Proportion Interval Estimation Methods*, Poster. NESUG 2001, Baltimore.
- [24] Wichura M. J. (2001): *Edgeworth Expansion*, skripta, Chicago. <http://galton.uchicago.edu/~wichura/>,Chicago.
- [25] Wilson E. B.,(1927): Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association* **22**, 209–212.